PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 :
C12N 1/20, 7/00, 15/00
C12P 21/00

A1

(11) International Publication Number: WO 90/01540

(43) International Publication Date: 22 February 1990 (22.02.90)

(21) International Application Number: PCT/US89/03417

(22) International Filing Date: 9 August 1989 (09.08.89)

(30) Priority data:
231,224 11 August 1988 (11.08.88) US

(71) Applicant: CALIFORNIA BIOTECHNOLOGY INC. [US/US]; 2450 Bayshore Parkway, Mountain View, CA 94043 (US).

(72) Inventors: HILLIKER, Sandra ; 3883 Buchanan, No. 158, Riverside, CA 92503 (US). WHITE, R., Tyler ; 41600 Marigold Drive, Fremont, CA 94539 (US).

(74) Agents: MURPHY, Lisabeth, Feix et al.; Irell & Manella, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: METHOD FOR STABILIZING HETEROLOGOUS PROTEIN EXPRESSION AND VECTORS FOR USE THEREIN

(57) Abstract

The present invention provides a method for stabilizing heterologous protein expression in bacteria by using a 3' truncated chloramphenicol acetyltransferase (CAT) gene fused in frame with a gene encoding a heterologous protein. When expressed in a bacterial host, the resulting hybrid gene produces a fusion protein in recoverable yield. Cleavage sites separating the CAT and heterologous protein are also provided to facilitate isolation and purification of the desired heterologous protein. The invention further provides bacterial vectors containing the hybrid gene fusions for expression of the fusion protein comprising the desired heterologous protein.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic			SN	Senegal
CG	Congo	KR	Republic of Korea	SU	Soviet Union
CH	Switzerland	LI	Liechtenstein	TD	Chad
CM	Cameroon	LK	Sri Lanka	TG	Togo
DE	Germany, Federal Republic of	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

METHOD FOR STABILIZING HETEROLOGOUS PROTEIN EXPRESSION AND VECTORS FOR USE THEREIN

Technical Field of the Invention

The present invention relates generally to the field of biotechnology. More particularly, the invention relates to the fields of protein expression and recombinant DNA technology to improve the yield of poorly expressed mammalian polypeptides in bacterial hosts.

Background of the Invention

Many eukaryotic proteins are not capable of being expressed in Escherichia coli in any measurable yield, or even if detectable, are not capable of being expressed at such commercially recoverable levels due to proteolysis of the foreign protein by the host. Small proteins (e.g., peptide hormones of less than 100 amino acids) appear to be especially sensitive to degradation. The degree of proteolysis varies from host to host and protein to protein. Possibly the highest level of expression of a eukaryotic protein in E. coli has been observed with gamma interferon, which was expressed at approximately 60% of total cellular protein. The high level of expression of a few eukaryotic proteins has been achieved because they reach a concentration in the cell where they can aggregate into insoluble masses called inclusion or refractile bodies (e.g., bovine growth hormone; Schoner et al (1985), Biotechnology 3:151-154). In this form, the eukaryotic protein is less susceptible to proteolysis.

Proteins which do not become insoluble on their own do in some cases form inclusion bodies if joined to another protein such as a procaryotic protein. A small number of prokaryotic proteins have been used in this manner: E. coli lacZ, trpE, and recA genes and the lambda cII gene, for example.

Chloramphenicol acetyltransferase (CAT) has been used as a selectable marker (resistance to chloramphenicol), as an easily assayed enzyme to monitor the efficiency of both eukaryotic and prokaryotic expression from different promoters (Delegeane, A.M., et al (1987) Mol Cell Biol 7:3994-4002), regulatory sequences, and/or ribosome binding sites, and for gene fusions which join sequences encoding a eukaryotic protein to the nucleotide sequence encoding mature, native CAT (Buckley and Hayashi (1986) Mol Gen Genet 204:120-125; European Patent Publication 161,937, published 21 November 1985) or to the carboxy terminal fragment of CAT (usually retaining CAT activity).

While the literature establishes that fusion proteins are useful to express heterologous proteins in bacteria and that the native CAT gene sequence has been used for such a purpose, efforts to use a truncated form of CAT to express or to increase the recoverable yield of heterologous, mammalian proteins such as amyloid protein A4-751 insert sequence, glucagon-like peptide I, adipsin/D, and lung surfactant SP-B and SP-C, have not been reported. In light of the fact that many important proteins cannot be successfully expressed in bacteria in any commercially recoverable yield, there is a need to develop systems for the bacterial expression and recovery of such proteins.

Disclosure of the Invention

One aspect of the invention concerns a method of stabilizing heterologous protein expression in a prokaryotic host comprising:

(a) constructing a hybrid gene comprising in sequential order, a 3' truncated chloramphenicol acetyltransferase (CAT) gene sequence fused in frame with a heterologous gene sequence encoding a mammalian polypeptide selected from the group consisting of amyloid protein A4-751 insert sequence, glucagon-like peptide I, adipsin/D, lung surfactant protein SP-B and lung surfactant protein SP-C; wherein said polypeptide is normally not recoverable in bacterial expression systems, and wherein said hybrid gene, upon translation, produces a fusion protein in a recoverable yield;

(b) providing a vector for expression of said hybrid gene;

(c) culturing the prokaryotic host transformed with the expression vector; and

(d) recovering the fusion protein.

A second aspect of the invention concerns a bacterial expression vector capable of enhancing the level of expression of non-stable, bacterially produced heterologous polypeptides comprising a hybrid gene having, in sequential order, a 3' CAT truncated gene sequence fused in frame to a heterologous gene sequence encoding a mammalian polypeptide selected from the group consisting of amyloid protein A4-751 insert sequence, glucagon-like peptide I, adipsin/D, lung surfactant protein SP-B and lung surfactant protein SP-C, wherein said polypeptide is normally not recoverable in bacterial expression systems; whereby said truncated CAT gene sequence is capable of rendering the resulting fusion protein resistant to proteolytic degradation.

A preferred embodiment for both the method and vector of the present invention employs a CAT coding

sequence of less than or equal to 180 amino acids, preferably between 73 and 180 amino acids. Although the resulting CAT protein is substantially reduced as compared to the native CAT protein, surprisingly, it has been found that the truncated CAT protein substantially contributes to the stability of the expressed protein and therefore, permits recovery of an increased yield of the desired heterologous protein.

Yet another aspect of the invention provides an improved bacterial expression vector capable of enhancing the level of expression of non-stable, bacterially produced heterologous polypeptides wherein said vector contains a hybrid gene having in sequential order, a modified 3' truncated CAT gene sequence linked to a heterologous gene sequence. The improvement comprises altering one or more DNA codons of the truncated CAT gene to eliminate potential chemical cleavage sites within the CAT protein.

Other aspects of the invention will be readily apparent to those of skill in the art from the description and examples which follow.

Brief Description of the Drawings

Figure 1 sets forth the amino acid and corresponding nucleotide sequences for a 241 amino acid (aa) CAT-hANP hybrid protein containing an endoproteinase Glu-C proteolytic cleavage site. The amino terminal portion of this hybrid protein encodes the first 210 amino acids of CAT, which sequence is extensively referred to throughout the present invention.

Figure 2 illustrates a series of vectors and synthetic fragments used for cloning and expression of the CAT-hANF hybrid proteins of the invention. Figure 2A depicts an EcoRI-PstI synthetic fragment containing the E. coli trp promoter-operator sequence, a ribosomal binding site, and downstream cloning sites. Figure 2B is a

restriction site and function map of plasmid pTrp233. Figure 2C is a restriction site and function map of plasmid pCAT21. Figure 2D is an EcoRI-HindIII synthetic fragment encoding the hANP (102-126) gene preceded by an endoproteinase Glu-C cleavage site. Figures 2E through G are restriction site and function maps of plasmids phNF75, pChNF109, and pChNF121, respectively. Figure 2H depicts a synthetic 1-73 aa CAT gene sequence contained within NdeI-HindIII fragment. Figure 2I is a restriction site and function map of plasmid pChNF142 wherein site-specific mutagenesis was used to substitute Tyr and Ser codons for residues 16 and 31, respectively, of the CAT gene.

Figure 3 illustrates two different preparative SDS-polyacrylamide gels. Figure 3A is an SDS-polyacrylamide gel of the CAT-A4-751i hybrid protein. Lane 1 = molecular size standards; Lane 2 = induced W3110 (pCAPi132); Lane 3 = induced W3110 (pTrp83) vector control; Lane 4 = uninduced W3110 (pCAPi136); and Lane 5 = induced W3110 (pCAPi136). Figure 3B is an SDS-polyacrylamide gel of the CAT-GLP-I hybrid protein. Lane 1 = molecular size standard; Lane 2 = uninduced W3110 (pCGLP139); Lane 3 = induced W3110 (pCGLP139); and Lane 4 = induced W3110 (pTrp83) vector control.

Figure 4 illustrates the amino acid and corresponding nucleotide sequences for a CAT-A4-751i hybrid protein and a CAT-GLP-I hybrid protein of the invention. Figure 4A depicts the first 73 codons encoding the amino terminus of the CAT protein joined in-frame to the synthetic A4-751i gene preceded by a chemical cleavage and site encoded by Asn-Gly. Figure 4B depicts the first 73 codons encoding the amino terminus of the CAT protein joined in-frame to the synthetic GLP-1 gene preceded by a Met codon.

Figure 5 illustrates two plasmids, pCAT73 and pCAT210, in which the gene for tetracycline resistance is restored in these CAT expression vectors.

Figure 6 is the nucleotide sequence and corresponding amino acid sequence of the SP-B expression construct pC210SP-B from the EcoRI site preceding the trp promoter region through the HindIII site containing the translation stop codon. The CAT, linker, and SP-B regions are identified therein, respectively, by the arrows.

Figure 7 is a preparative SDS-polyacrylamide gel of the CAT:SP-B fusion protein. Lane A = molecular size standards; Lane B = induced W3110 cells containing pTrp233 vector control; and Lane C = induced pC210SP-B/W3110 cells.

Figure 8 illustrates the nucleotide sequence and corresponding amino acid sequence of the 251 residue CAT:SP-C fusion protein from plasmid pC210SP-C. The CAT gene, linker sequence and SP-B gene are sequentially identified therein by the arrows.

Figure 9 provides the molecular weight determinations for each of the CAT:SP-C fusion proteins. Lane A = molecular size standards; Lane B = induced W3110 cells containing pTrp233 vector control; Lane C = induced pC106SP-C; Lane D = pC149SP-C; Lane E = pC179SP-C; and Lane F = pC210SP-C.

Figure 10 provides the cDNA and amino acid sequences for human adipsin/D.

Modes for Carrying Out the Invention

A. Definitions

As used herein the term "stabilizing protein expression" refers to a property of a fusion protein responsible for inhibiting proteolysis of a foreign protein by a recombinant host cell.

"Insoluble" as referred to proteins intends a condition wherein a protein may be recovered only by extraction with detergents or chaotropic agents. Usually,

insoluble proteins are formed as a consequence of intracellular aggregation of the cloned gene products.

"High protein expression" or "enhanced protein expression" refers to a level of expression wherein the fused protein can comprise 10% or more of the total protein produced by each cell. A preferred range for high protein expression levels is from 10-20% of total cell protein.

As used herein, "non-recoverable" refers to a level of expression wherein the desired protein may be detected using sensitive techniques, e.g., Western blot analysis, yet the protein is not commercially recoverable using conventional purification techniques such as SDS-polyacrylamide gel electrophoresis, gel filtration, ion exchange chromatography, hydrophobic chromatography, affinity chromatography, or isoelectric focusing.

"Mammalian" refers to any mammalian species, and includes rabbits, mice, dogs, cats, primates and humans, preferably humans.

As used herein, the term "heterologous" proteins refers to proteins which are foreign to the host cell transformed to produce them. Thus, the host cell does not generally produce such proteins on its own.

B. CAT Fusions

CAT encodes a 219 amino acid mature protein and the gene contains a number of convenient restriction endonuclease sites (5'-PvuII, EcoRI, DdeI, NcoI, and ScaI-3') throughout its length to test gene fusions for high level expression. These restriction sites may be used for ease of convenience in constructing the hybrid gene sequences of the invention or other sites within the gene sequence may be generated using techniques commonly known to those of skill in the art. Any of the resulting CAT sequences are considered useful so long as the resulting

CAT fusion retains the ability to enhance the expression of the desired heterologous peptide.

The expression constructs of the invention can employ most of the CAT-encoding gene sequence or a substantially truncated portion of the sequence encoding an N-terminal portion of the CAT protein linked to the gene encoding the desired heterologous polypeptide. In one embodiment of the invention, the CAT portion of the fusion codes for about the N-terminal one-third of the CAT sequence.

The expression constructs exemplified herein, which demonstrated enhanced levels of expression for a variety of heterologous proteins, utilize a number of varying lengths of the CAT protein ranging in size from 73 to 210 amino acids. The 73 amino acid CAT fusion component is conveniently formed by digesting the CAT nucleotide sequence at the EcoRI restriction site. Similarly, the 210 amino acid CAT fusion component is formed by digesting the CAT nucleotide sequence with ScaI. These, as well as other CAT restriction fragments, may then be ligated to any nucleotide sequence encoding a desired protein to enhance expression of the desired protein.

Significantly, although the expression level of fusion protein (approximately 15-20% of total cell protein) was similar for the CAT (106 amino acid) - SP-C fusion and the CAT (210 amino acid) - SP-C fusion, it can be seen that the former case actually represents a significant increase in expression level for the desired SP-C polypeptide, since the SP-C polypeptide constitutes a substantially larger proportion of the total fusion protein in the former case. The ability to increase expression level for the desired polypeptide by reducing the size of the fused CAT protein sequence was quite an unexpected finding in view of the experience of the prior art. In general, the prior art experience has been that

reduction in size of the bacterial leader sequence does not result in increased production of the fused heterologous polypeptide due to a concomitant larger reduction in the expression level of the fusion protein.

With one exception, the various CAT-heterologous fusion proteins exemplified herein were found to be expressed in the range of approximately 10-20% of the total cell protein. Thus, the versatility of the CAT fusions, that is, the ability to use a variety of CAT coding sequences having the ability to enhance the expression of a desired protein, allows great flexibility of choice when constructing CAT hybrid genes.

The reading frame for translating the nucleotide sequence into a protein begins with a portion of the amino terminus of CAT, the length of which varies, continuing in-frame with or without a linker sequence into the protein to be expressed, and terminating at the carboxy terminus of the protein. An enzymatic or chemical cleavage site may be introduced downstream of the CAT sequence to permit recovery of the cleaved product from the hybrid protein. Such cleavage sequences are known in the art as are the conditions under which cleavage can be effected. Following cleavage, the desired heterologous polypeptide can be recovered using known techniques of protein purification. Suitable cleavage sequences include, without limitation, cleavage following methionine residues (cyanogen bromide), glutamic acid residues (endoproteinase Glu-C), tryptophan residues (N-chlorosuccinimide with urea or with sodium dodecyl sulfate (SDS)) and cleavage between asparagine and lysine residues (hydroxylamine).

To avoid internal cleavage within the CAT sequence, amino acid substitutions can be made using conventional site specific mutagenesis techniques (Zoller, M.J., and Smith, M. (1982), Nuc Acids Res 10:6487-6500, and Adelman, J.P., et al (1983), DNA 2:183-193). This is conducted using a synthetic oligonucleotide primer com-

plementary to a single-stranded phage DNA to be mutagenized except for limited mismatching, representing the desired mutation. Of course, these substitutions would only be performed when expression of CAT is not significantly affected. Where there is only one internal cysteine residue, as in the short CAT sequence, this residue may be replaced to help reduce multimerization through disulfide bridges.

C. CAT Fusion Vectors

Procaryotic systems may be used to express the CAT fusion sequence; procaryotic hosts are, of course, the most convenient for cloning procedures. Procaryotes most frequently are represented by various strains of E. coli; however, other microbial strains may also be used. Plasmid vectors which contain replication sites, selectable markers and control sequences derived from a species compatible with the host are used; for example, E. coli is typically transformed using derivatives of pBR322, a plasmid derived from an E. coli species by Bolivar et al, Gene 2:95 (1977). pBR322 contains genes for ampicillin and tetracycline resistance, and thus provides multiple selectable markers which can be either retained or destroyed in constructing the desired vector.

In addition to the modifications described above which would facilitate cleavage and purification of the product polypeptide, the gene conferring tetracycline resistance may be restored to the exemplified CAT fusion vectors for an alternative method of plasmid selection and maintenance.

Although the E. coli tryptophan promoter-operator sequences have been exemplified in the present CAT vectors, different control sequences can be substituted for the trp regulatory sequences and are considered to be within the scope of the invention. Commonly used procaryotic control sequences which are defined

herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequence, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (lac) promoter systems (Chang et al, Nature 198:1056), the lambda-derived P_L promoter (Shimatake et al, Nature 292:128 (1981)) and N-gene ribosome binding site, and the trp-lac (trc) promoter system (Amann and Brosius, Gene 40:183 (1985)).

Since the general utility of these CAT vectors have been established with very different mammalian peptides (ranging in protein size, the presence or absence of disulfide bonds, and being hydrophobic or hydrophilic in nature) vectors with unique restriction sites may be created or substituted for the pBR322-derived vector illustrated in the examples.

D. Heterologous Protein Expression

Amino terminal DNA sequences of CAT have been fused to DNA sequences encoding human polypeptides for high level expression in the bacterial host E. coli. The polypeptides described herein are relatively small mammalian polypeptides ranging in size from about 30 to 76 amino acid residues. Attempts to directly express, e.g., in a non-fused form, each of these polypeptides in bacteria have been unsuccessful, most likely due to the proteolytic degradation which occurs upon translation of the mRNA product. In the case of extremely hydrophobic polypeptides, even attempts to express such polypeptides using beta-galactosidase fusions produced detectable but very low level amounts of protein.

Examples of polypeptides that have been successfully expressed to high level in bacteria using the truncated CAT fusions include a variety of mammalian polypeptides including amyloid protein A4-751 insert sequence, glucagon-like peptide I, adipsin/D, lung surfactant protein SP5 (SP-C), and lung surfactant SP18

(SP-B). Preferably, the mammalian protein is of human origin, although other sources are also contemplated to be within the scope of this invention. A4-751 is a 57 amino acid sequence identified within the precursor for the A4 amyloid protein associated with Alzheimer's disease and shares homology with the Kunitz family of serine proteinase inhibitors (Ponte, P., et al (1988) Nature 331:525-527; Tanzi, R.E., et al (1988) Nature 331:528-530). Glucagon-like peptide I (GLP-I, 7-31) is a 31 amino acid hormone co-encoded in the glucagon gene which is a potent stimulator of insulin release (Mojsov, S., et al (1987) J Clin Inves 79:616-619). Adipsin/D is a serine protease synthesized in and secreted from adipocytes (Zusalak, K.M., et al (1985) J Mol Cell Biol 5:419). Lung surfactant SP-B is a 76 amino acid hydrophobic protein. Lung surfactant SP-C is a 35 amino acid hydrophobic protein. Both SP-B and SP-C greatly enhance spreading of surfactant phospholipids at an air:water interface.

E. Hosts Exemplified

Host strains used in cloning and procaryotic expression herein are as follows:

For cloning and sequencing, and for expression of construction under control of most bacterial promoters, E. coli strains such as MC1061, DH1, RR1, W3110, MM294, B, C600hfl, K803, HB101, JA221, and JM101 may be used.

F. General Methods

Recombinant DNA methods are described in Maniatis et al (1982), Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, when not specifically cited in the following examples. Methods are also described in the literature for visualizing inclusion bodies, isolating them from cells, then solubilizing, purifying, and cleaving the hybrid protein (e.g., Itakura, K., et al (1977) Science 198:1056-1063; Shine, J., et al

(1980) Nature 285:455-461). Methods are also available, if necessary, for refolding the protein product (Creighton, T.E., Proceedings of Genex-UCLA Symposium, 1985, Kingstones (in press). The teachings of all of these references are incorporated herein by reference.

Examples

I. Expression of Chloramphenicol Acetyltransferase-Human Atrial Natriuretic Peptide Hybrid Proteins in Escherichia coli.

A. Expression vector pChNF109.

Expression vector pChNF109 encodes a 241 amino acid CAT-hANP hybrid protein containing an endoproteinase Glu-C proteolytic cleavage site (Fig. 1). Most of the CAT gene (amino acids 1-210) has been joined in-frame to the hANP(102-126) gene and cleavage site (26 amino acids) through a linker sequence (5 amino acids). The hANP polypeptide comprises about 10% of the hybrid protein. This vector was constructed from plasmids pTrp233, pCAT21, and phNF75 which supplied the plasmid backbone and trp promoter-operator, the CAT gene, and the hANP(102-126) gene and cleavage site, respectively.

1. Construction of pChNF109.

Plasmid pTrp233 was constructed by insertion of a synthetic EcoRI-PstI fragment containing the E. coli trp promoter-operator sequence, a ribosomal binding site, and downstream cloning sites into plasmid pKK233-2-NdeI which contains strong transcription termination signals, T1T2, and the beta-lactamase gene. The synthetic fragment (see Fig. 2A) was assembled using the method of Vlasuk et al (1986), J. Biol Chem 261: 4789-4796 and its sequence confirmed by the method of Sanger et al (1977), Proc Natl Acad Sci USA 74:5463-5467 in M13mp8 and M13mp9. Plasmid

pKK233-2-NdeI (disclosed in co-pending U.S. Serial No. 766,030, filed 8 May 1985 and incorporated herein by reference) was digested with EcoRI and PstI, its termini dephosphorylated using calf intestinal phosphatase, and ligated with the synthetic EcoRI-PstI fragment. Plasmid pTrp233 was isolated (Fig. 2B) from E. coli JA221 transformed to ampicillin resistance.

Plasmid pCAT21 was constructed by insertion of the CAT gene (from transposon Tn9, Alton and Vapnek, (1979) Nature 282:864-869) into plasmid pTrp233 under the control of the trp promoter-operator. Plasmid pAL13ATCAT (a plasmid disclosed in co-pending U.S. Serial No. 095,742, filed 11 September 1987 and incorporated herein by reference) was digested with NdeI and HindIII and the approximately 750 bp NdeI-HindIII fragment containing the CAT gene (with the initiating Met residue encoded at the NdeI site) was purified using agarose gel electrophoresis. The CAT gene was ligated with NdeI and HindIII-digested pTrp233 using T4 DNA ligase. From E. coli MC1061 (Casadaban et al (1980), I Mol Biol 138: 179-209) ampicillin-resistant transformants, plasmid pCAT21 was isolated (Fig. 2C).

Plasmid phNF75 was constructed by insertion of a synthetic hANP gene preceded by a proteolytic cleavage site into plasmid pBgal (Shine et al (1980), Nature 285:456). Eight oligodeoxyribonucleotides (Fig. 2D) were assembled into a synthetic hANP(102-126) gene preceded by an endoproteinase Glu-C cleavage site (method of Vlasuk et al (1986), supra). The synthetic DNA fragment (with a 5' EcoRI tail and a 3' blunt end) was ligated with EcoRI and SmaI restriction endonuclease digested M13mp19 using T4 DNA ligase for the purpose of DNA sequencing (method of Sanger et al (1977), supra). A clone with the correct sequence, M13-hNF7, was digested with BamHI and BglII, the fragment containing the hANP gene purified by agarose gel electrophoresis, and the fragment ligated with BamHI-

digested and bacterial alkaline phosphatase dephosphorylated pTrp233 using T4 DNA ligase. A plasmid with the insert in the orientation which gives adjacent HindIII, BamHI and EcoRI sites at the 3' end of the hANP gene, phNF73, was identified by the size of the fragments generated by digestion with HindIII and PvuII. Plasmid phNF73 was digested with EcoRI, the hANP gene purified using polyacrylamide gel electrophoresis, and the gene ligated with EcoRI-digested and bacterial alkaline phosphatase dephosphorylated plasmid pBgal. From E. coli MC1061 ampicillin-resistant transformants, plasmid phNF75 (Fig. 2E) was identified by the size of the DNA fragments generated by digestion with PstI and HindIII.

Expression vector pChNF109 was constructed by insertion of DNA fragments containing CAT, hANP and the proteolytic cleavage site, and a linker sequence into plasmid pTrp233. Plasmid phNF75 was digested with EcoRI and HindIII, the approximately 80 bp EcoRI-HindIII fragment containing hANP was purified by polyacrylamide gel electrophoresis, and ligated with EcoRI- and HindIII-digested pTrp233 using T4 DNA ligase. From E. coli MC1061 ampicillin-resistant transformants, plasmid phNF87 was isolated and digested with BamHI and the fragments were dephosphorylated using bacterial alkaline phosphatase. A BamHI cassette containing the trp promoter-operator, ribosomal binding site, and large amino terminal fragment of the CAT gene was generated by digesting pCAT21 with ScaI, attaching BamHI synthetic linkers (5'-CGGATCCG-3') to the blunt termini using T4 DNA ligase, digesting the ligation with BamHI and purification of the approximately 740 bp BamHI fragment by agarose gel electrophoresis. The BamHI cassette and plasmid phNF87 were ligated using T4 ligase and ampicillin-resistant transformants of E. coli MC161 obtained. Plasmid pChNF109 (Fig. 2F), with the BamHI cassette in the orientation such that the CAT gene is fused in-frame to the endoproteinase Glu-C cleavage

site followed by the hANP gene, was selected on the basis of DNA fragment size in an EcoRI digest of the plasmid.

2. Expression of CAT(1-210)-hANP(102-126) Hybrid Protein From Plasmid pChNF109.

Plasmid pChNF109 expresses a CAT-hANP(102-126) hybrid protein under the control of the E. coli trp promoter-operator. The plasmid was used to transform E. coli W3110 (ATCC Accession No. 27325) to ampicillin resistance and one colony was grown in culture overnight at 37°C in complete M9 medium containing M9 salts, 2 mM MgSO4, 0.1 mM $CaCl_2$, 0.4% glucose, 0.5% casamino acids, 40 ug/ml tryptophan, 2 ug/ml thiamine hydrochloride, and 100 ug/ml ampicillin sulfate. The overnight culture was diluted 100-fold into the same M9 medium described above (uninduced culture) and into M9 medium in which the tryptophan had been replaced by 25 ug/ml of 3-beta-indoleacrylic acid (induced culture).

Expression was assessed after shaking the cultures for 6 hr at 37°C. The uninduced culture had reached a high cell density (stationary phase) and the induced culture was still at a low cell density (exponential phase). Phase-contrast microscopy revealed cells of normal morphology in the uninduced culture and elongated cells containing several refractile inclusion bodies in the induced culture. Total cell protein samples were prepared by boiling cell pellets in Laemmli buffer for 5 min and were analyzed by electrophoresis through a 12% SDS-polyacrylamide gel followed by staining of the protein with Coomassie Blue.

B. Expression Vector pChNF121.

Expression vector pChNF121 encodes a 99 amino acid CAT-hANP hybrid protein containing an endoproteinase Glu-C proteolytic cleavage site (Fig. 4A). Approximately one-third of the CAT gene (amino acids 1-73) has been

fused to the hANP(102-126) gene and proteolytic cleavage site (26 amino acids) without an intervening linker. The hANP polypeptide comprises 25% of the hybrid protein. This vector was constructed from plasmids pChNF109 and phNF87 which supplied the amino terminal fragment of the CAT gene and the hANP gene and proteolytic cleavage site, respectively.

1. Construction of pChNF121.

Plasmid phNF87 was digested with EcoRI, its termini dephosphorylated with bacterial alkaline phosphatase, and ligated with an approximately 320 bp EcoRI fragment containing the trp promoter-operator, ribosome binding site, and amino-terminus of the CAT gene. This EcoRI cassette was purified from an EcoRI digest of PChNF109 using agarose gel electrophoresis. Plasmid pChNF121 (Fig. 2G) was isolated from the ampicillin-resistant transformants of E. coli MC1061. On the basis of the size of the DNA fragments from a PvuII digest of the plasmid, the CAT and hANP genes were inferred to be fused in-frame to produce a hybrid protein.

2. Expression of CAT(1-73)-hANP(102-126) Hybrid Protein From Plasmid pChNF121.

Plasmid pChNF121 expresses a CAT-hANP(102-126) hybrid protein under the control of the E. coli trp promoter-operator. The plasmid was used to transform E. coli W3110 (prototroph, TrpR+) to ampicillin resistance and one colony was grown in culture overnight at 37°C in complete M9 medium (see Section A.2.). The overnight culture was diluted 100-fold into complete M9 medium (uninduced culture) and into M9 medium with 25 ug/ml 3-beta-indole-acrylic acid replacing the 40 ug/ml tryptophan (induced culture).

Expression was assessed after shaking the cultures for 6 hr at 37°C. The uninduced culture had

reached a high cell density whereas the induced culture reached about one-third this density. Phase contrast microscopy revealed cells of normal morphology in the uninduced culture and elongated cells with several refractile inclusion bodies in the induced culture. Total cell protein samples were prepared by boiling cell pellets in Laemmli buffer for 5 min. and were analyzed by electrophoresis through a 12% SDS-polyacrylamide gel followed by staining of the protein with Coomassie Blue.

C. Expression Vector pChNF142.

Expression vector pChNF142 encodes a 99 amino acid CAT-hANP hybrid protein containing a unique Trp residue following amino acid residue 73 of the CAT protein, as a site for chemical cleavage. Approximately one-third of the CAT gene (amino acids 1-73) has been fused to the hANP(102-126) gene and chemical cleavage site (26 amino acids). This amino terminal fragment of CAT has been modified to substitute a Tyr residue for Trp[16] and a Ser residue for Cys[31] to remove the additional chemical cleavage site and reduce the multimerization of the hybrid protein through disulfide bridges. A synthetic hANP gene preceded by sequence encoding a Trp residue has been assembled for this vector.

1. Construction of pChNF142.

Plasmid pTrp233 was digested with EcoRI, its termini filled in with E. coli DNA polymerase I, Klenow fragment, and ligated with T4 DNA ligase (to remove the EcoRI restriction endonuclease cleavage site). From ampicillin-resistant transformants of E. coli MC1061, plasmid pTrp81 was isolated and shown to resist cleavage by EcoRI. Plasmid pTrp81 was digested with NdeI and HindIII, purified by agarose gel electrophoresis, and ligated with a synthetic CAT gene fragment using T4 DNA ligase. The synthetic NdeI-HindIII CAT gene fragment

(Fig. 2H) was assembled from three pairs of oligodeoxyribonucleotides as previously described. From ampicillin-resistant transformants of E. coli MC1061, plasmid pCAT127 was isolated and shown to contain the synthetic CAT fragment by digestion with EcoRI and AvaI. The plasmid was digested with BamHI and HindIII, the BamHI-HindIII fragment containing CAT was purified by agarose gel electrophoresis, sequenced by the method of Sanger et al (1977), supra, and the correct DNA sequence confirmed.

Plasmid pCAT127 was digested with EcoRI and HindIII and ligated using T4 DNA ligase with a pair of annealed synthetic oligodeoxyribonucleotides encoding hANP(102-126) preceded by a Trp residue on an EcoRI-HindIII DNA fragment. Plasmid pChNF142 (Fig. 2I) was isolated from ampicillin-resistant transformants of E. coli MC1061. Insertion of the hANP gene was confirmed by the size of the DNA fragments in a BamHI and HindIII digest of the plasmid. The sequence of the hANP gene was confirmed from an EcoRI-ScaI agarose gel purified fragment from pChNF142.

2. Expression of CAT(1-73), Tyr[16] Ser[31]-hANP(102-126) pChNF142.

The expression of a modified CAT-hANP(102-126) hybrid protein is conducted in substantial accordance with the teaching of the previous examples A.2 and B.2.

II. Expression of Chloramphenicol Acetyltransferase--Amyloid A4 Protein Insert (A4-751i) Hybrid Proteins in Escherichia coli.

In the following examples high level expression of the 57 amino acid insert within the amyloid A4-751 protein was achieved by fusing a synthetic A4-751i gene to DNA sequences encoding amino terminal fragments of CAT under the control of the E. coli tryptophan promoter-

operator on a pBR322-derived plasmid. The synthetic A4-751i gene encodes amino acids 289-345 from amyloid A4-751 protein (Ponte et al (1988), Nature 331:525-527) preceded by a chemical cleavage site, Asn-Gly. Hydroxylamine cleavage of the hybrid protein between these two residues will yield the insert protein with a Gly residue at its amino terminus.

A. Expression Vector pCAPi132.

Expression vector pCAPi132 encodes a 132 amino acid CAT-A4751i hybrid protein containing a hydroxylamine cleavage site (Fig. 4A). Approximately the amino terminal third of the CAT gene (amino acids 1-73) has been joined in-frame to the A4-751i gene and cleavage site (59 amino acids). The A4-751i protein comprises about 43% of the hybrid protein. This vector was constructed from plasmids pTrp233 and pChNF121 and the synthetic A4-751i gene and cleavage site.

1. Construction of pCAPi132.

Plasmid pTrp233 was digested with EcoRI and HindIII, purified by agarose gel electrophoresis, and ligated with the synthetic gene encoding the A4-751i protein and cleavage site using T4 DNA ligase. The gene had been assembled from six oligodeoxyribonucleotides using previously described techniques and its sequence (Fig. 4A) confirmed. Plasmid pAPi131 was isolated from ampicillin-resistant transformants of E. coli MC1061. Insertion of the synthetic gene was confirmed by the size of the DNA fragments from a PvuI and BamHI digest of plasmid mini-prep DNA.

Plasmid pAPi131 was digested with EcoRI to linearize the vector and its termini dephosphorylated using bacterial alkaline phosphatase. Plasmid pChNF121 was digested with EcoRI and the approximately 320 bp EcoRI fragment containing the trp promoter-operator, ribosome

binding site, and amino terminus of the CAT gene (amino acids 1-73) was purified by agarose gel electrophoresis. This EcoRI cassette was ligated with the pAPi131 plasmid using T4 DNA ligase and ampicillin-resistant transformants of MC1061 were obtained. On the basis of DNA fragment size in a PvuII digest of mini-prep plasmid DNA, plasmid pCAPi132 was isolated with an in-frame fusion of CAT and A4-751i sequences.

2. Expression of CAT(1-73)-A4-751i Hybrid Protein From Plasmid pCAPi132.

Plasmid pCAPi132 expresses a CAT-A4-751i hybrid protein under the control of the E. coli trp promoter-operator. The plasmid was used to transform E. coli W3110 to ampicillin resistance and one colony was grown in culture overnight at 37°C in complete M9 medium. The overnight culture was diluted 100-fold into complete M9 medium which contains 40 ug/ml tryptophan (uninduced culture) and into complete M9 medium containing 25 ug/ml 3-beta-indoleacrylic acid instead of tryptophan (induced culture).

Expression was assessed after shaking the cultures for 6 hr at 37°C. The uninduced culture had reached a high cell density whereas the induced culture was at a lower cell density. Phase contrast microscopy revealed cells of normal morphology in the uninduced culture and cells with "pre-inclusion bodies" in the induced culture. As used herein, "pre-inclusion bodies" are defined as less refractile bodies which appear to convert in time to the more refractile "inclusion bodies" as the hybrid protein accumulates in the cells. Total cell protein samples were prepared by boiling cell pellets in Laemmli buffer for 5 min and then analyzed by electrophoresis through a 12% SDS-polyacrylamide gel followed by staining with Coomassie Blue (Fig. 3A). This CAT(1-73)-A4-751i hybrid protein migrates between the

lysozyme (14,300 MW) and beta-lactoglobulin (18,400 MW) protein standards on this gel. Using a Kontes fiber optic scanner and Hewlett-Packard Integrator to scan the gel, the hybrid protein was estimated to comprise about 7% of the total cell protein. This is a moderate expression level of E. coli but A4-751i comprises almost half of the hybrid protein.

To confirm the presence of A4-751i in the hybrid protein, Western blot analysis was carried out on an unstained 12% SDS-polyacrylamide gel of these protein samples. Protein was blotted to nitrocellulose and incubated with anti-A4-751i serum (prepared against a 16 amino acid synthetic peptide containing amino acids 11-26 of the 57 amino acid insert protein). After incubation with ^{125}I-protein A (Amersham) the blot was placed on X-ray film at -70°C for several days. The synthetic peptide anti-serum detected the hybrid protein as well as several other E. coli proteins.

B. Expression Vector pCAPi136.

Expression vector pCAPi136 encodes a 274 amino acid CAT-A4-751i hybrid protein containing a hydroxylamine cleavage site. Most of the CAT gene (amino acids 1-210) has been joined in-frame to the A4-751i gene and cleavage site (59 amino acids) through a linker sequence (5 amino acids). The A4-751i polypeptide comprises about 21% of the hybrid protein. This vector was constructed from plasmids pAPi131 and pChNF109.

1. Construction of pCAPi136.

Plasmid pAPi131 was digested with EcoRI to linearize the vector and its termini dephosphorylated using bacterial alkaline phosphatase. From a partial EcoRI digest of pChNF109 an approximately 740 bp EcoRI fragment containing the trp promoter-operator, the CAT gene (amino acids 1-210), and linker sequence was purified

by agarose gel electrophoresis. This EcoRI cassette and vector pAPi131 were ligated using T4 DNA ligase and ampicillin-resistant transformants of E. coli MC1061 were isolated. From the size of DNA fragments in plasmid mini-preps digested with BamHI, plasmid pCAPi136 was isolated with the CAT gene and the synthetic A4-751i gene in-frame.

2. Expression of CAT(1-210)-A4-751i Hybrid Protein From Plasmid pCAPi136.

Plasmid pCAPi136 expresses a CAT-A4-751i hybrid protein under the control of the E. coli trp promoter-operator. The plasmid was used to transform E. coli W3110 to ampicillin resistance and one colony was grown in culture overnight at 37°C in complete M9 medium. The overnight culture was diluted 100-fold into the same M9 medium (uninduced culture) and into M9 complete medium containing 25 ug/ml 3-beta-indoleacrylic acid instead of tryptophan (induced culture).

Expression was assessed after shaking the cultures for 6 hr at 37°C. Both the uninduced and induced cultures reached high cell densities. Phase contrast microscopy revealed cells of normal morphology in the uninduced cultures and cells containing inclusion bodies or pre-inclusion bodies (50:50) in the induced cultures. Total cell protein samples were prepared by boiling cell pellets in Laemmli buffer for 5 min and were analyzed by electrophoresis through a 12% SDS-polyacrylamide gel followed by staining with Coomassie Blue (Fig. 3A). This CAT-A4-751i hybrid protein migrates between the alpha-chymotrypsinogen (25,700 MW) and ovalbumin (43,000 MW) protein standards on this gel. Using a Kontes fiber optic scanner and Hewlett-Packard Integrator to scan the gel, the hybrid protein was estimated to comprises about 15% of total cell protein. This is moderately high level expression for E. coli.

To confirm the presence of A4-75li in the hybrid protein, Western blot analysis was carried out on an unstained 12% SDS-polyacrylamide gel of these protein samples. Using the method described above (section II. A.2.), the synthetic peptide anti-serum detected the hybrid protein as well as several other E. coli proteins.

III. Expression of Chloramphenicol Acetyltransferase--Glucagon-Like Peptide I (7-37) Hybrid Protein in Escherichia coli.

In the following example, high level expression of the 31 amino acid GLP-I(7-37) was achieved by fusing a synthetic GLP-I gene to DNA sequences encoding an amino terminal fragment of CAT under the control of the E. coli tryptophan promoter-operator on a pBR322-derived plasmid. The synthetic gene encodes amino acids 7-37 of GLP-1 (Mojsov et al (1987), J. Clin Invest 79:616-619) preceded by a Met residue. Treatment with cyanogen bromide releases the insulinotropic peptide.

A. Expression Vector pCGLP139.

Expression vector pCGLP139 encodes a 105 amino acid CAT-GLP-I hybrid protein containing a cyanogen bromide cleavage site (Fig. 4B). Approximately the amino terminal third of the CAT gene (amino acids 1-73) has been joined in-frame to the GLP-1 gene and cleavage site (32 amino acids). The GLP-I peptide comprises about 30% of the hybrid protein. This vector was constructed from plasmids pTrp233 and pChNF109 and the synthetic GLP-I gene and cleavage site.

1. Construction of pCGLP139.

Plasmid pTrp233 was digested with EcoRI and HindIII, purified by agarose gel electrophoresis, and ligated with the synthetic gene using T4 DNA ligase. The gene had been assembled from four oligodeoxyribo-

nucleotides and its sequence (Fig. 4B) confirmed. From ampicillin-resistant transformants of E. coli MC1061, plasmid pGLP138 was isolated. Insertion of the synthetic gene was confirmed by the failure of plasmid mini-prep DNA to be cut by PstI.

Plasmid pGLP138 was digested with EcoRI to linearize the vector, its termini dephosphorylated using bacterial alkaline phosphatase, and ligated with the EcoRI cassette from plasmid pChNF109 using T4 DNA ligase. Plasmid pChNF109 had been digested with EcoRI and the approximately 320 bp EcoRI fragment containing the trp promoter-operator, ribosome binding site, and an amino terminal fragment of the CAT gene purified by agarose gel electrophoresis. Plasmid pCGLP139 was isolated from ampicillin-resistant transformants of MC1061. On the basis of DNA fragment size in an AvaI and PvuII digest of plasmid mini-prep DNA, the fusion of CAT and GLP-I sequences was confirmed to be in-frame.

2. Expression of CAT(1-73)-GLP-I(7-37) Hybrid Protein From Plasmid pCGLP139.

Plasmid pCGLP139 expresses a CAT-GLP-I hybrid protein under the control of the E. coli trp promoter-operator. The plasmid was used to transform E. coli W3110 to ampicillin resistance and one colony was grown in culture overnight at 37°C in complete M9 medium. The overnight culture was diluted 100-fold into complete M9 medium which contains 40 ug/ml tryptophan (uninduced culture) and into complete M9 medium in which 25 ug/ml 3-beta-indoleacrylic acid has been substituted for the tryptophan (induced culture).

Expression was assessed after shaking the cultures for 6 hr at 37°C. The uninduced culture had reached a high cell density whereas the induced culture was at a lower cell density. Phase contrast microscopy revealed cells of normal morphology in the uninduced

culture and elongated cells with three or more refractile inclusion bodies in the induced culture. Total cell protein samples were prepared by boiling cell pellets in Laemmli buffer for 5 min and were analyzed by electrophoresis through a 12% SDS-polyacrylamide gel followed by staining with Coomassie Blue (Fig. 3B). This CAT(1-73)-GLP-I(7-37) hybrid protein migrates between the bovine trypsin inhibitor (6,200 MW) and lysozyme (14,300 MW) protein standards. Using a Kontes fiber optic scanner and Hewlett-Packard Integrator to scan the gel, the hybrid protein was estimated to comprise about 20% of the total cell protein. (Considering the number of inclusion bodies observed per cell, all of the hybrid protein may not have been solubilized in the Laemmli buffer, and this estimate may be low.) This is high level expression for E. coli.

The molecular weight of the hybrid protein is as predicted for this gene fusion. Amino acid composition analysis of the purified hybrid protein or protein sequencing of the peptide after cyanogen bromide cleavage can be performed to confirm its expression.

IV. CAT Fusion With Human SP-B and SP-C.

The mature forms of both human SP-C and SP-B are expressed as fusions with portions of bacterial CAT. The surfactant peptides are joined to the carboxy terminus of the CAT sequences through a hydroxylamine-sensitive asparagine-glycine linkage. The CAT-surfactant fusions are expressed from the tryptophan promoter of the bacterial vector pTrp233.

A. Expression Vector pC210SP-B.

SP-B expression vector pC210SP-B encodes a fusion protein of 293 residues in which 210 amino acids of CAT are joined to the 76 amino acids of SP-B through a linker of 7 amino acids containing the hydroxylamine-sensitive cleavage site. Cleavage of the fusion with

hydroxylamine releases a 77 amino acid SP-B product containing the 76 residue mature form of SP-B, plus an amino-terminal glycine residue.

To construct pC210SP-B, the short EcoRI-HindIII segment containing ANF sequences was removed from pChNF109, and replaced by a portion of human SP-B cDNA #3 extending from the PstI site at nucleotide (nt) 643 (Fig. 6) to the SphI site at nt 804. The EcoRI site was joined at the PstI site through two complementary oligonucleotides encoding the hydroxylamine sensitive cleavage site as well as the amino-terminal residues of mature SP-B (oligo #2307: 5'-AAT TCA ACG GTT TCC CCA TTC CTC TCC CCT ATT GCT GGC TCT GCA-3' and oligo #2308: 5'-GAC CCA GCA ATA GGG GAG AGG AAT GGG GAA ACC GTT G-3'). The SphI site was joined to the HindIII site of PTrp233 through a second set of complementary nucleotides encoding the carboxy-terminal residues of mature SP-B (oligo #3313: 5'-AGC TTA CCG GAG GAC GAG GCG GCA GAC CAG CTG GGG CAG CAT G-3' and oligo #3314: 5'-CTG CCC CAG CTG GTC TGC CGC CTC GTC CTC CGG TA-3').

The expression plasmid was used to transform E. coli stain W3110 to ampicillin resistance. Rapidly growing cultures of pC210SP-B/W3110 in M9 medium were made 25 ug/ml IAA (3-beta indoleacrylate, Sigma I-1625) to induce the Trp promoter. By 1 hr after induction, refractile cytoplasmic inclusion bodies were seen by phase contrast microscopy inside the still-growing cells. 5 hr after induction, the equivalent of 1 $O.D._{550}$ of cells were pelleted by centrifugation, then boiled for 5 min in SDS sample buffer for electrophoresis in a 12% SDS-polyacrylamide gel followed by staining with Coomassie Blue (Fig. 7). Lane A = molecular size standards; Lane B = induced W3110 cells containing pTrp233 vector control; and Lane C = induced pC210SP-B/W3110. The predicted molecular weight of the CAT:SP-B fusion protein is 45,000 daltons. The hybrid CAT:SP-B protein was estimated to

comprise 15-20% of the total cell protein in the induced cultures.

B. CAT Fusions with SP-C.

A series of vectors were constructed encoding fusion proteins in which mature human SP-C was fused to the carboxy termini of different portions of CAT through a hydroxylamine-sensitive asparagine-glycine linkage. Hydroxylamine cleavage of the fusion protein produced by each construct releases a mature SP-C of 35 amino acids which lacks the amino-terminal phenylalanine residue seen in a portion of natural human SP-C.

1. pC210SP-C.

The amino acid sequence of the 251 residue fusion protein encoded plasmid pC210SP-C. The 210 amino acids of CAT are joined to 35 amino acids of mature SP-C through a linker of 6 amino acids. The mature SP-C portion of the total fusion protein comprises 14%.

In Fig. 8 is shown the nucleotide sequence of pC210SP-C, in which the EcoRI-HindIII fragment of pC210SP-B containing SP-B sequences has been replaced by a segment of human SP-C cDNA #18 extending from the ApaLI site at nucleotide 123 to the AvaII site at nucleotide 161. The EcoRI site of the CAT vector was joined to the SP5 ApaLI site through two complementary oligonucleotides encoding the hydroxylamine sensitive cleavage site as well as the amino-terminal residues of mature SP-C (oligo #2462: 5'-AAT TCA ACG GCA TTC CCT GCT GCC CAG-3' and oligo #2463: 5'-TGC ACT GGG CAG CAG GGA ATG CCG TTG-3'). The AvaII site of SP-C was joined to the HindIII site of pC210SP-B through a second set of complementary nucleotides encoding the carboxy-terminal residues of mature SP-C and a stop codon (oligo #2871: 5'-AGC TTA GTG GAG ACC CAT GAG CAG GGC TCC CAC AAT CAC CAC GAC GAT GAG-3' and oligo #2872: 5'-GTC

CTC ATC GTC GTG GTG ATT GTG GGA GCC CTG CTC ATG GGT CTC CAC TA-3').

2. pC179SP-C.

The amino acid sequence of the 217 residue fusion protein encoded by pC179SP-C is a slight modification of the sequence shown in Fig. 8. In pC179SP-C, the 179 amino acids of CAT are joined to 35 amino acids of mature SP-C through a linker of 3 amino acids (Glu, Phe, Asn). SP-C portion of the total fusion protein comprises 16%.

To construct pC179SP-C, a portion of the CAT sequence was removed from pC210SP-C. Starting with pC210SP-C, a DNA fragment extending from the NcoI site at nt 603 (Fig. 8) to the EcoRI site at nt 728 was removed, and the NcoI and EcoRI cohesive ends were rejoined with two complementary oligonucleotides (oligo #3083: 5'-CAT GGG CAA ATA TTA TAC GCA AG-3' and oligo #3084: 5'-AAT TCT TGC GTA TAA TAT TTG CC-3'). In effect, 31 residues of CAT, and 3 residues of the linker polypeptide are missing in the new fusion protein encoded by vector pC179SP-C.

3. pC149SP-C.

The amino acid sequence of the 187 residue fusion protein encoded by pC149SP-C is a slight modification of the sequence shown in Fig. 8. In plasmid pC149SP-C, the 149 amino acids of CAT are joined to 35 amino acids of mature SP-C through a linker of 3 amino acids (Glu, Phe, Asn). The SP-C portion of the total fusion protein comprises 18.7%.

To construct pC149SP-C, a portion of the CAT segment of pC210SP-C extending from the DdeI site at nt 523 (Fig. 8) to the EcoRI site at nt 728 was removed and replaced by a set of two complementary oligonucleotides (oligo #3082: 5'-TCA GCC AAT CCC G-3' oligo #3081: 5'-AAT TCG GGA TTG GC-3').

4. pC106SP-C.

The amino acid sequence of the 144 residue fusion protein encoded by pC106SP-C is a slight modification of the sequence shown in Fig. 8. In plasmid pC106SP-C, the 106 amino acids of CAT are joined to 35 amino acids of mature SP-C through a linker of 3 amino acids (Glu, Phe, Asn). The SP-C portion of the total fusion protein comprises 24%.

pC106SP-C was constructed by replacing the EcoRI fragment of pC210SP-C (nt 302 to nt 728, Fig. 8) with two sets of complementary oligos which were annealed, then ligated together through a region of homology (oligo #3079: 5'-AAT TCC GTA TGG CAA TGA AAG ACG GTG AGC TGG TGA TAT GGG ATA GTG TTC ACC CTT GT-3' was annealed with oligo #3085: 5'-ACA CTA TCC CAT ATC ACC AGC TCA CCG TCT TTC ATT GCC ATA CGG-3'; oligo #3080: 5'-TAC ACC GTT TTC CAT GAG CAA ACT GAA ACG TTT TCA TCG CTC TGG G-3' was annealed with oligo #3078: 5'-AAT TCC CAG AGC GAT GAA AAC GTT TCA GTT TGC TCA TGG AAA ACG GTG TAA CAA GGG TGA-3').

5. Expression From SP-C Vectors.

Each SP-C expression vector was used to transform E. coli strain W3110 to ampicillin resistance. Rapidly growing cultures of expression strains were induced as described above. By 1 hr after induction, refractile cytoplasmic inclusion bodies were seen by phase contrast microscopy inside the still-growing cells. 5 hr after induction, the equivalent of 1 $O.D._{550}$ of cells were pelleted by centrifugation, then boiled for 5 min in SDS sample buffer for electrophoresis in a 12% SDS-polyacrylamide gel followed by staining with Coomassie Blue. The results are provided in Fig. 9 wherein Lane A = molecular size standards, Lane B = induced W3110 cells containing pTrp233 vector control; Lane C = induced pC106SP-C; Lane D = pC149SP-C; Lane E = pC179SP-C; Lane F = pC210SP-C. The hybrid CAT:SP-C protein produced by each

vector is estimated to comprise 15-20% of the total cell protein in the induced cultures.

V. Improved CAT Vectors for Expression of Hybrid Proteins in Escherichia Coli.

In the following examples, the basic CAT gene fusion vector has been improved in several ways: (1) unique cloning sites are created for insertion of the gene to be expressed, (2) the CAT gene is modified to optimize cleavage and/or purification of the peptides, and (3) the gene conferring resistance to tetracycline is restored to provide an alternative method for plasmid selection and maintenance.

A. Expression Vectors pCAT73 and pCAT210.

Expression vector pCAT73 contains genes conferring resistance to both ampicillin and tetracycline, unique EcoRI and HindIII cloning sites for insertion of genes to be expressed, and the amino terminal fragment (1-73) of the CAT gene. The cleavage site, included with the inserted gene, may not be unique. This plasmid is constructed from plasmids pBR322, pTrp233, pCAT21, and oligodeoxyribonucleotides. Expression vector pCAT210 differs from pCAT73 in that it contains the larger amino terminal fragment (1-210) of the CAT gene from which the EcoRI site at the sequence encoding residues 72 and 73 (Glu-Phe) has been removed. (An alternative codon choice preserves the Glu and permits the use of unique EcoRI and HindIII cloning sites.) Other DNA fragments encoding the amino terminus of the CAT gene, smaller than 73 amino acids or between 73 and 210 amino acids may also be constructed by insertion of an EcoRI site at the desired fusion point.

1. Construction of pCAT73.

Restoration of the gene for tetracycline resistance requires restoring the BamHI-HindIII-EcoRI fragment of pBR322 to the CAT expression vector. Since the unique cloning sites desired for this vector are EcoRI and HindIII, this must be done in a manner which removes these sites but retains resistance to tetracycline. Since insertion of DNA at the HindIII site upstream of the coding region often prevents gene expression, this site is removed by creating a point mutation at the HindIII site. Plasmid pBR322, was digested with EcoRI and HindIII and the vector backbone gel purified. The backbone was ligated with synthetic EcoRI-HindII fragments, which are formed by annealing pairs of oligonucleotides using T4 DNA ligase. The fragments contain the normal EcoRI-HindIII sequence with the exception of point mutations (G or C) at the first adenine of the recognition sequence 5'-AAGCTT-3'. An intermediate plasmid was isolated from ampicillin-resistant and tetracycline-resistant E. coli MC1061 transformants whose plasmid mini-prep DNA was not digested by HindIII.

A BamHI-EcoRI fragment no longer containing a HindIII site was purified from agarose gel electrophoresis from a BamHI and EcoRI digest of plasmid pTetH1. The fragment was ligated using T4 DNA ligase with plasmid pTrp233 which was also digested with BamHI and EcoRI and agarose gel purified. Transformed with the ligation, colonies of E. coli MC1061 were selected for ampicillin and/or tetracycline resistance. Plasmid pTrpT233 was resistant to both antibiotics.

In an alternate embodiment, digestion of pTrpT233 with EcoRI, blunting of the termini with DNA polymerase I, Klenow fragment, and ligation with T4 DNA ligase will eliminate the EcoRI site (which does not affect resistance to tetracycline). Tetracycline-resistant plasmid pTrpT234 which has lost undesirable

HindIII and EcoRI sites is isolated from colonies of E. coli MC1061 transformed with this ligation.

The CAT gene is obtained as an NdeI-HindIII fragment purified by agarose gel electrophoresis of an NdeI-HindIII digest of pCAT21. Plasmid pTrpdeltaHind was digested with NdeI and HindIII, purified by agarose gel electrophoresis, and ligated with the CAT gene using T4 DNA ligase. From ampicillin (or tetracycline) resistant transformants of E. coli MC1061 digested with EcoRI and HindIII to verify incorporation of the CAT gene, plasmid pCAT73 (Fig. 5A) is isolated.

2. Construction of pCAT210.

The BamHI-HindIII fragment containing the trp promoter-operator, ribosome binding site, and CAT gene is purified by agarose gel electrophoresis from a BamHI and HindIII digest of plasmid pCAT21. Site specific mutagenesis is carried out on the fragment using M13 and mutagenic oligodeoxyribonucleotides to convert the GAA codon for Glu to GAG (also to Glu) within the EcoRI site, 5'-GAATTC-3'. One such plasmid, M13-CATdR, is digested with ScaI to linearize the vector and ligated with an EcoRI linker (for the same reading frame as in pCAT73) using T4 DNA ligase. From the transfectants, M13-CATR1, is isolated and digested with NdeI and HindIII. The new CAT gene is purified by agarose gel electrophoresis and ligated using T4 DNA ligase with NdeI-HindIII-digested plasmid pTrpT234. Plasmid pCAT210 (Fig. 5B) is isolated from ampicillin (or tetracycline) resistant transformants of E. coli MC1061.

B. Expression Vectors pCAT73-T and pCAT73-M.

Expression vectors pCAT73-T and pCAT73-M are examples in which the amino acid sequence of CAT has been altered using site specific mutagenesis techniques to facilitate purification of the product protein. In these

cases, the Trp residue at position 16 may be substituted with Tyr and the Met residue at position 67 may be substituted by Ile or Leu to eliminate potential chemical cleavage sites within CAT. In addition, the Cys at position 31 may also be substituted using a conservative amino acid alteration, that is, substitution with an amino acid which does not adversely affect biological activity. Preferred residues include alanine, serine, leucine, isoleucine and valine, most preferred is serine. These latter alterations are intended to reduce multimerization through disulfide bridges.

C. Expression of Modified CAT-GLP-1

Plasmid pTrpdeltaHind contains the restored Tet^R gene from pTrp233 (although the HindIII site has been eliminated), the Trp_{16} to Tyr, Cys_{31} to Ser, and Met_{67} to Leu substitutions in the CAT gene sequence, and the GLP-1 gene (taught in Example III) fused in-frame to the modified CAT gene through a methione residue. The vector was used to transform several E. coli strains including W3110, MC1061, DH1, MM294 and RR1.

E. coli RR1 transformants were more stable and appeared to have better induction/repression control of the Trp promoter than any of the other hosts. An alternative construction for this vector includes reversing the Tet^R gene (to avoid the back-to-back placement of the Tet^R and Trp promoters in the present construct) to alleviate the stability problems observed using bacterial hosts other than RRI transformants.

VI. Construction of pTrpCAT72:Adipsin/D.

The coding sequence for mature human adipsin/D was fused to pCAT72 to produce a fusion protein suitable, for example, to generate antisera against human adipsin/D.

A. Construction of pTrpCAT72 Q3S1

Plasmid pCAT72 Q3S1 was constructed to eliminate Asn residues at which secondary cleavages can occur during hydroxylamine release of peptides fused to CAT. The Asn residues at amino acid positions 26, 51 and 78 of CAT were changed to Gln residues. At the same time, the single Cys at position 31 was changed to Ser to decrease the amount of aggregation seen with many CAT fusion proteins.

The vector pCAT72 Q3S1 was constructed as follows: Oligos CAT72-1 through 6 (below) were annealed and ligated into pUC-9 which had been cleaved with NdeI and EcoRI. In this way, the mutated CAT72 was joined to the polylinker region of the pUC plasmid. CAT72 Q3S1 with the polylinker was then removed from pUC by cleavage with NdeI and HindIII, and inserted into pTrp233 between NdeI and HindIII to yield pTrpCAT72 Q3S1.

CAT72-1

```
        10         20         30         40         50
TATGGAGAAA AAAATCACTG GATATACCAC CGTTGATATA TCCCAATGGC

        60         70
ATCGTAAAGA ACATTTTGAG GCATTTCA
```

CAT72-2

```
        10         20         30         40         50
CAAAATGTTC TTTACGATGC CATTGGGATA TATCAACGGT GGTATATCCA

        60
TGATTTTTT TCTCCA
```

CAT72-3

```
        10         20         30         40         50
TCAGTTGCT CAATCTACCT ATCAGCAGAC CGTTCAGCTG GATATTACGG

        60         70         80
CCTTTTTAAA GACCGTAAAG AAACAGAAGC
```

CAT72-4

```
        10         20         30         40         50
CTTTACGGTC TTTAAAAAGG CCGTAATATC CAGCTGAACG GTCTGCTGAT

        60         70         80
AGGTAGATTG AGCAACTGAC TGAAATGCCT
```

CAT72-5

```
          10          20          30          40          50
ACAAGTTTTA TCCGGCCTTT ATTCACATTC TTGCCCGCCT GATGCAGGCT

CATCCGG
```

CAT72-6

```
          10          20          30          40          50
AATTCCGGAT GAGCCTGCAT CAGGCGGGCA AGAATGTGAA TAAAGGCCGG

          60          70
ATAAAACTTG TGCTTCTGTT T
```

B. Construction of pTrpCAT72 Q6S3

Starting with pCAT72 Q3S1, pCAT153 Q6S3 was constructed to change the Asn residues at positions 130, 141 and 148 of CAT to Gln residues, and to change the Cys residues at 91 and 126 to Ser residues.

Plasmid CAT72 Q3S1 in pUC-9 was cleaved with EcoRI. Oligos CAT153-1 through 6 (below) were annealed and ligated into pCAT72 to give pCAT153 Q6S3. The modified pCAT153 was then removed from pUC by cleavage with NdeI and HindIII, and the resulting fragment inserted into pTrp233 to give pTrpCAT153 Q6S3.

CAT153-1

```
          10          20          30          40          50
AATTTCGTAT GGCAATGAAA GACGGTGAGC TGGTGATATG GGATAGTGTT

          60          70          80
CACCCTTCTT ACACCGTTTT CCATGAGCAA
```

CAT153-2

```
          10          20          30          40          50
AAAACGGTGT AAGAAGGGTG AACACTATCC CATATCACCA GCTCACCGTC

          60
TTTCATTGCC ATACGA
```

CAT153-3

```
          10          20          30          40          50
ACTGAAACGT TTTCATCGCT CTGGAGTGAA TACCACGACG ATTTCCGGCA

          60          70          80
GTTTCTACAC ATATATTCGC AAGATGTGGC
```

CAT153-4

```
         10          20          30          40          50
GCGAATATAT GTGTAGAAAC TGCCGGAAAT CGTCGTGGTA TTCACTCCAG

         60          70          80
AGCGATGAAA ACGTTTCAGT TTGCTCATGG
```

CAT153-5

```
         10          20          30          40          50
GTCTTACGGT GAACAGCTGG CCTATTTCCC TAAAGGGTTT ATTGAGCAGA

         60          70
TGTTTTTCGT CTCAGCCCAG CCCG
```

CAT153-6

```
         10          20          30          40          50
AATTCGGGCT GGGCTGAGAC GAAAAACATC TGCTCAATAA ACCCTTTAGG

         60          70          80
GAAATAGGCC AGCTGTTCAC CGTAAGACGC CACATCTT
```

Next, the human adipsin/D cDNA hg31-40 (Figure 10) was constructed. The BamHI-StyI fragment containing the mature coding region was gel purified and inserted into pUC-9 which had been cleaved with BamHI and HindIII. The StyI end of the cDNA was joined to the HindIII end of pUC using two oligos (#3886 5'-CATGGGTGCCGGGGCCTGA-3' and #3887 5'-AGCTTCAGGCCCCGGCACC-3'). By inserting the BamHI-StyI fragment of adipsin/D into pUC in this way, the coding sequence of adipsin/D was placed in frame with the EcoRI site of pUC-9. The EcoRI-HindIII fragment of this construct was removed from pUC-9 and inserted into pTrpCAT72 between the EcoRI site and the HindIII sites to yield pTrpCAT72:Adipsin/D.

This construct gave 10-15% levels of fusion protein upon induction in W3110 cells.

Modifications of the above described modes for carrying out the invention that are obvious to those of skill in the art of molecular biology, protein chemistry, cell biology, or related fields are intended to be within the scope of the following claims.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A method of stabilizing heterologous protein expression in a prokaryotic host comprising:

(a) constructing a hybrid gene comprising in sequential order, a 3' truncated chloramphenicol acetyltransferase (CAT) gene sequence fused in frame with a heterologous gene sequence encoding a mammalian polypeptide selected from the group consisting of amyloid protein A4-751 insert sequence, glucagon-like peptide I, adipsin/D, lung surfactant protein SP-B and lung surfactant protein SP-C, wherein said polypeptide is normally not recoverable in bacterial expression systems, and wherein said hybrid gene, upon translation, produces a fusion protein in a recoverable yield;

(b) providing a vector for expression of said hybrid gene;

(c) culturing the prokaryotic host transformed with the expression vector; and

(d) recovering the fusion protein.

2. The method of claim 1 wherein said prokaryotic host is a bacterial cell.

3. The method of claim 2 wherein said bacterial cell is E. coli.

4. The method of claim 1 wherein said 3' truncated CAT gene sequence enhances the level of heterologous protein present in the total cellular protein.

5. The method of claim 1 wherein the length of the truncated CAT gene sequence encodes a CAT peptide of about 73 to about 210 amino acids.

6. The method of claims 1 or 5 wherein said hybrid gene further comprises a DNA sequence encoding a selective cleavage site located between the CAT gene sequence and the heterologous gene sequence.

7. The method of claim 6 wherein said selective cleavage site is composed of tryptophan, methionine, asparagine-glycine, or glutamic acid.

8. A method of stabilizing heterologous protein expression in a prokaryotic host comprising:

(a) constructing a hybrid gene comprising in sequential order, a 3' truncated chloramphenicol acetyltransferase (CAT) gene sequence encoding a CAT peptide of about 73 to about 180 amino acids, fused in-frame with a heterologous gene sequence encoding a mammalian polypeptide selected from the group consisting of amyloid protein A4-751 insert sequence, glucagon-like peptide I, adipsin/D, lung surfactant protein SP-B and lung surfactant protein SP-C, wherein said heterologous protein is normally not recoverable in bacterial expression systems, and wherein said hybrid gene, upon translation, produces a fusion protein in a recoverable yield;

(b) providing a vector for expression of said hybrid gene;

(c) culturing the prokaryotic host transformed with the expression vector; and

(d) recovering the fusion protein.

9. The method of claim 8 wherein said hybrid gene further comprises a DNA sequence encoding a selective

cleavage site located between the CAT gene sequence and the heterologous gene sequence.

10. A bacterial expression vector capable of enhancing the level of expression of non-stable, bacterially produced heterologous polypeptides comprising:

a hybrid gene having in sequential order, a 3' truncated CAT gene sequence linked to a heterologous gene sequence encoding a mammalian polypeptide selected from the group consisting of amyloid protein A4-751 insert sequence, glucagon-like peptide I, adipsin/D, lung surfactant protein SP-B and lung surfactant protein SP-C, wherein said polypeptide is normally not recoverable in bacterial expression systems, whereby said truncated CAT gene sequence is capable of rendering the resulting fusion protein resistant to proteolytic degradation.

11. The method of claim 10 wherein the length of the truncated CAT gene sequence encodes a CAT peptide of about 73 to about 210 amino acids.

12. The bacterial expression vector of claims 10 or 11 wherein said hybrid gene further comprises a DNA sequence encoding a selective cleavage site located between the CAT gene sequence and the heterologous gene sequence.

13. The vector of claim 12 wherein the hybrid gene having said 3' truncated CAT gene sequence, upon expression, enhances the level of the heterologous protein present in the total cellular protein.

14. In a bacterial expression vector capable of enhancing the level of expression of non-stable, bacterially produced heterologous polypeptides wherein the vector comprises a hybrid gene having in sequential order, a 3'

truncated CAT gene sequence linked to a heterologous gene sequence encoding a polypeptide normally not recoverable in bacterial expression systems, said truncated CAT gene sequence being capable of rendering the resulting fusion protein resistant to proteolytic degradation, the improvement comprising altering one or more DNA codons of the truncated CAT gene to eliminate potential chemical cleavage sites within the CAT protein.

15. The improved bacterial expression vector of claim 32 wherein the alterations include substituting the DNA encoding a) methionine at position 67 of CAT with DNA encoding isoleucine or leucine; (b) cysteine at position 31 of CAT with DNA encoding serine; or (c) tryptophan at position 16 of CAT with DNA encoding tyrosine.

FIG. 1-1

```
                                                                  10                  20
NH2-MET Glu Lys Lys Ile Thr Gly Tyr Thr Thr Val Asp Ile Ser Gln Trp His Arg Lys Glu
    ATG GAG AAA AAA ATC ACT GGA TAT ACC ACC GTT GAT ATA TCC CAA TGG CAT CGT AAA GAA

                                          30                  40
His Phe Glu Ala Phe Gln Ser Val Ala Gln Cys Thr Tyr Asn Gln Thr Val Gln Leu Asp
CAT TTT GAG GCA TTT CAG TCA GTT GCT CAA TGT ACC TAT AAC CAG ACC GTT CAG CTG GAT

                                                    50                  60
Ile Thr Ala Phe Leu Lys Thr Val Lys Lys Asn Lys His Lys Phe Tyr Pro Ala Phe Ile
ATT ACG GCC TTT TTA AAG ACC GTA AAG AAA AAT AAG CAC AAG TTT TAT CCG GCC TTT ATT

                                                              70                  80
His Ile Leu Ala Arg Leu MET Asn Ala His Pro Glu Phe Arg MET Ala MET Lys Asp Gly
CAC ATT CTT GCC CGC CTG ATG AAT GCT CAT CCG GAA TTC CGT ATG GCA ATG AAA GAC GGT

                                                                        90                  100
Glu Leu Val Ile Trp Asp Ser Val His Pro Cys Tyr Thr Val Phe His Glu Gln Thr Glu
GAG CTG GTG ATA TGG GAT AGT GTT CAC CCT TGT TAC ACC GTT TTC CAT GAG CAA ACT GAA

                                                                  110                 120
Thr Phe Ser Ser Leu Trp Ser Glu Tyr His Asp Asp Phe Arg Gln Phe Leu His Ile Tyr
ACG TTT TCA TCG CTC TGG AGT GAA TAC CAC GAC GAT TTC CGG CAG TTT CTA CAC ATA TAT

                                                            130                 140
Ser Gln Asp Val Ala Cys Tyr Gly Glu Asn Leu Ala Tyr Phe Pro Lys Gly Phe Ile Glu
TCG CAA GAT GTG GCG TGT TAC GGT GAA AAC CTG GCC TAT TTC CCT AAA GGG TTT ATT GAG

                                                                  150                 160
Asn MET Phe Phe Val Ser Ala Asn Pro Trp Val Ser Phe Thr Ser Phe Asp Leu Asn Val
AAT ATG TTT TTC GTC TCA GCC AAT CCC TGG GTG AGT TTC ACC AGT TTT GAT TTA AAC GTG

                                                                        170                 180
Ala Asn MET Asp Asn Phe Phe Ala Pro Val Phe Thr MET Gly Lys Tyr Tyr Thr Gln Gly
GCC AAT ATG GAC AAC TTC TTC GCC CCC GTT TTC ACC ATG GGC AAA TAT TAT ACG CAA GGC
```

CAT

CAT

Asp Lys Val Leu MET Pro Leu Ala Ile Gln Val His His Ala Val Cys Asp Gly Phe His
GAC AAG GTG CTG ATG CCG CTG GCG ATT CAG GTT CAT CAT GCC GTT TGT GAT GGC TTC CAT

190 200

hANP (102-126)

Val Gly Arg MET Leu Asn Glu Leu Gln Gln Ser Asp Pro Glu Phe Glu
GTC GGC AGA ATG CTT AAT GAA TTA CAA CAG TCG GAT CCG GAA TTC GAA

210

Arg Ser Ser Cys Phe Gly Gly Arg MET Asp Arg Ile Gly Ala Gln Ser Gly Leu Gly Cys
CGC TCT TCT TGT TTC GGT CGT ATG GAT CGT ATC GGT GCT CAA TCT GGT TTG GGT TGT

220 230

Asn Ser Phe Arg Tyr-COOH
AAC TCT TTC AGA TAC

240

FIG. 1-2

```
                                                    NdeI  KpnI PstI
EcoRI          -35          -10
5' AGAATTCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGAACTAGTTAACTAGTACGCAAGTTCACGTAAAAAGGGTATCACATATGGTACCTGCAGA 3'
3' TCTTAAGTTTATAAGACTTTACTCGACAACTGTTAATTAGTAGCTTGATCAATTGATCATGCGTTCAAGTGCATTTTTCCCATAGTGTATACCATGGACGTCT 5'
        tryptophan promoter-operator          mRNA start     S.D.  NH2-Met-.....
```

FIG. 2A

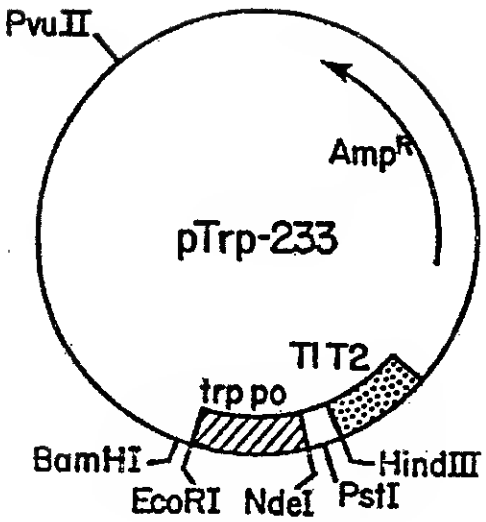
PvuII
Amp^R
pTrp-233
T1T2
trp po
BamHI
EcoRI
NdeI
PstI
HindIII

FIG. 2B

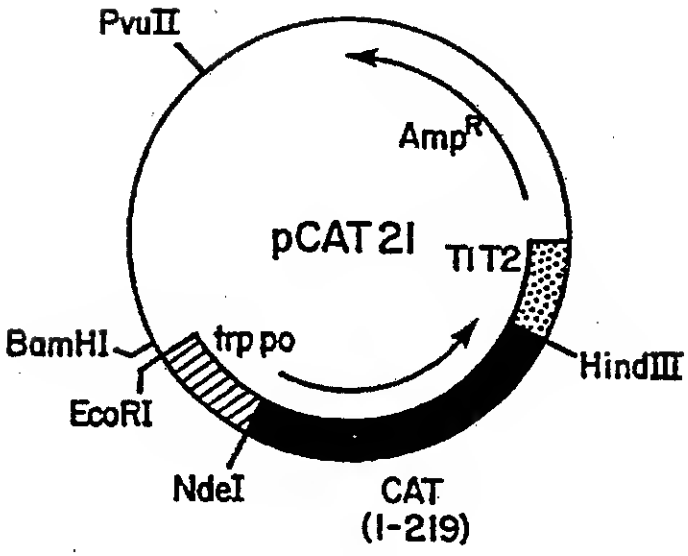
PvuII
Amp^R
pCAT 21
T1T2
BamHI
trp po
EcoRI
NdeI
CAT
(1-219)
HindIII

FIG. 2C

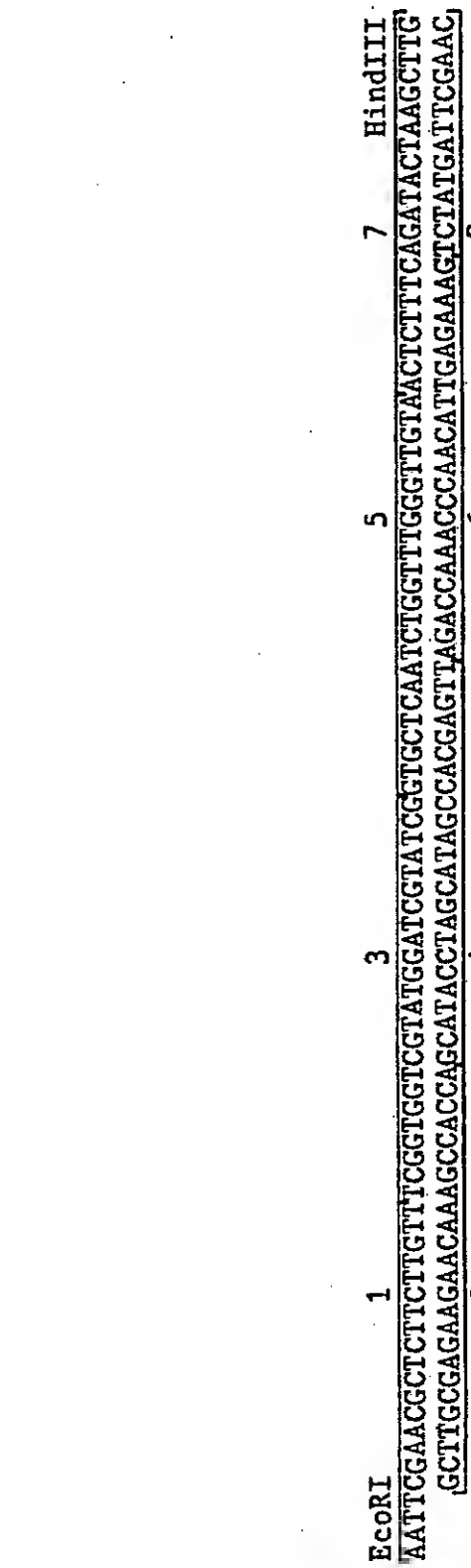
EcoRI
1
3
5
7
HindIII
AATTCGAACGCTCTTCTTGTTTCGGTGGTCGTATGGATCGTATCGGTGCTCAATCTGGTTTGGGTTGTAACTCTTTCAGATACTAAGCTTG
GCTTGCGAGAAGAACAAAGCCACCAGCATACCTAGCATAGCCACGAGTTAGACCAAACCCAACATTGAGAAAGTCTATGATTCGAAC
2
4
6
8

FIG. 2D

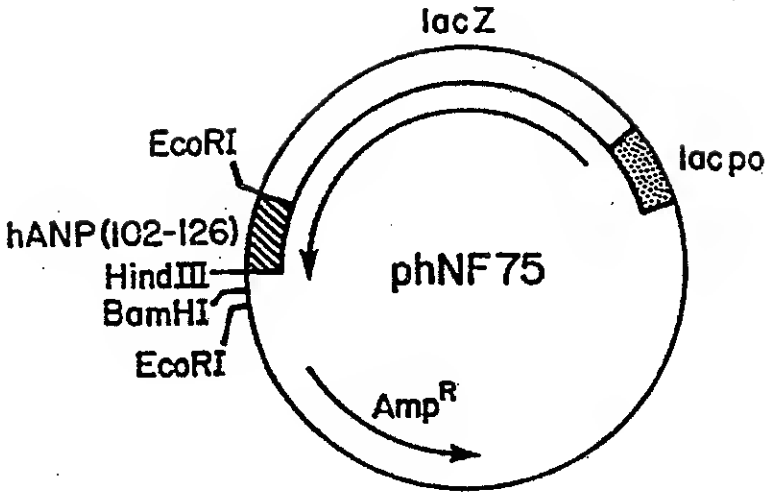
lacZ
EcoRI
lac po
hANP(102-126)
HindIII
BamHI
EcoRI
phNF75
Amp^R

FIG. 2E

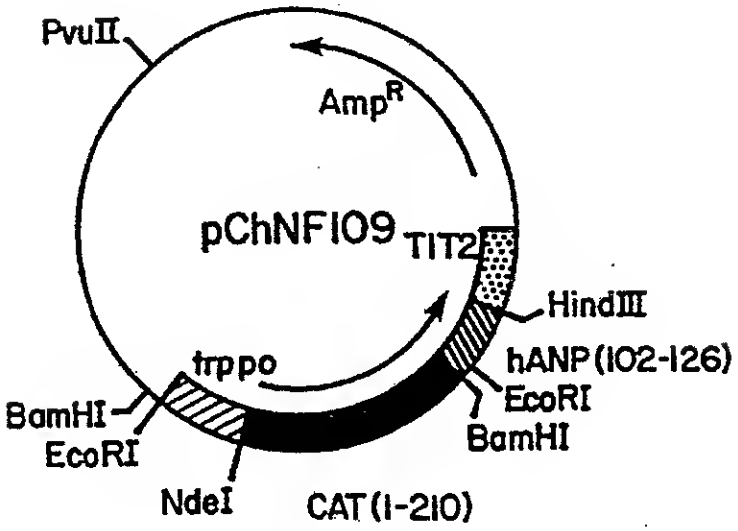
PvuII
Amp^R
pChNF109
TIT2
HindIII
hANP(102-126)
EcoRI
BamHI
trp po
BamHI
EcoRI
NdeI
CAT(1-210)

FIG. 2F

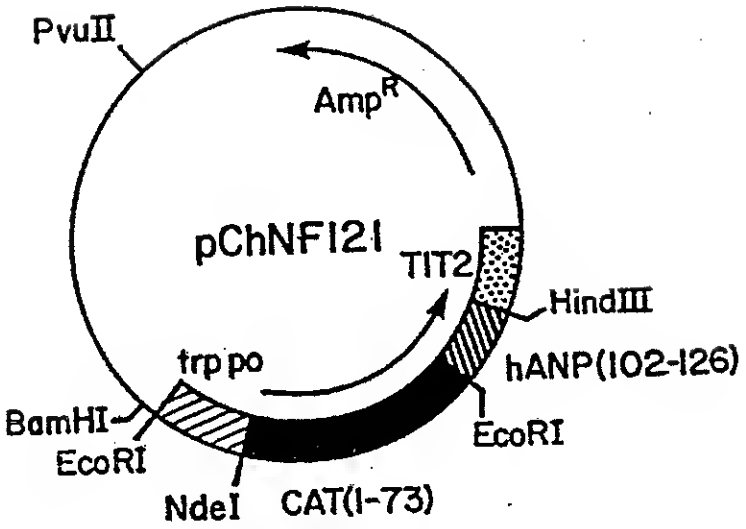
PvuII
Amp^R
pChNF121
TIT2
HindIII
hANP(102-126)
trp po
BamHI
EcoRI
EcoRI
NdeI
CAT(1-73)

FIG. 2G

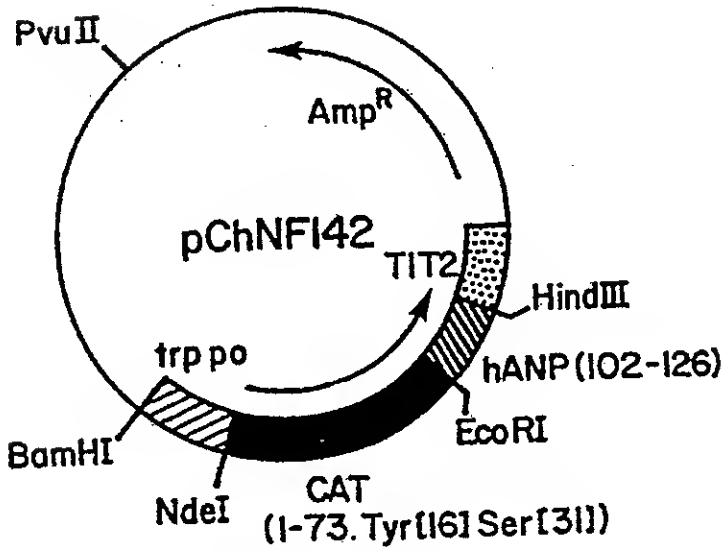
PvuII
Amp^R
pChNF142
TIT2
HindIII
trp po
hANP(102-126)
BamHI
EcoRI
NdeI
CAT
(1-73. Tyr[16] Ser[31])

FIG. 2H

```
                NdeI
(1)  5' TATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATATCATCGTAAAGAACATTTT 3'
(2)     3' ACCTCTTTTTTTAGTGACCTATATATGGTGGCAACTATATAGGGTTATAGTAGCATTTCTTGTAAAACTCCGT 5'


(3)  5' GAGGCATTTCAGTCAGTTGCTCAATCAACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACC 3'
(4)     3' AAAGTCAGTCAACGAGTTAGTTGGATATTGGTCTGGCAAGTCGACCTATAATGCCGGAAAAATTTCTGGCATTTC 5'


                                                  EcoRI      HindIII
(5)  5' GTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCATTTA 3'
(6)     3' TTTTTATTTCGTGTTCAAAATAGGCCGGAAATAAGTGTAAGAACGGGCGGACTACTTACGAGTAGGCCTTAAGTAAATTCGA 5'
```

FIG. 21

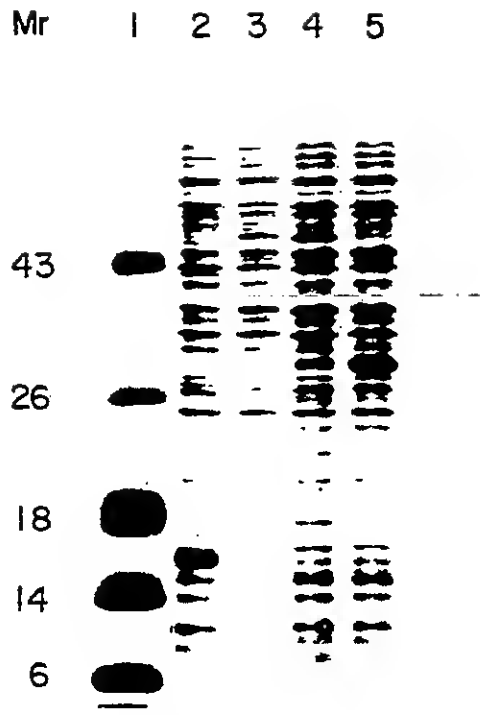

FIG. 3A

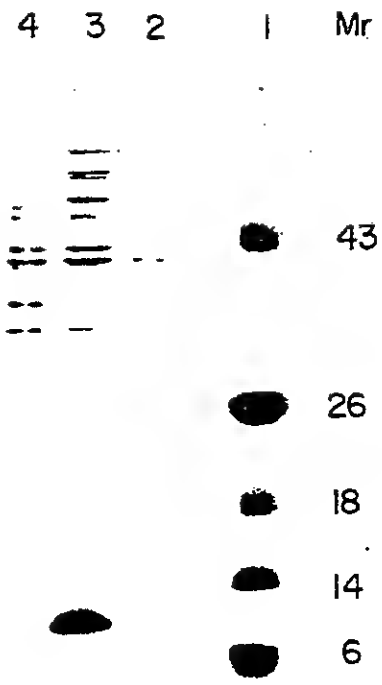

FIG. 3B

FIG. 4A

CAT

```
                                       10                                      20
NH2-MET Glu Lys Lys Ile Thr Cly Tyr Thr Thr Val Asp Ile Ser Gln Trp His Arg Lys Clu
    ATG CAG AAA AAA ATC ACT CGA TAT ACC ACC CTT CAT ATA TCC CAA TCC CAT CCT AAA GAA

                                       30                                      40
    His Phe Glu Ala Phe Gln Ser Val Ala Cln Cys Thr Tyr Asn Gln Thr Val Gln Leu Asp
    CAT TTT CAG CCA TTT CAG TCA GTT CCT CAA TCT ACC TAT AAC CAG ACC GTT CAG CTG GAT

                                       50                                      60
    Ile Thr Ala Phe Leu Lys Thr Val Lys Lys Asn Lys His Lys Phe Tyr Pro Ala Phe Ile
    ATT ACG GCC TTT TTA AAC ACC GTA AAC AAA AAT AAG CAC AAC TTT TAT CCG CCC TTT ATT

                                       70
    His Ile Leu Ala Arg Leu MET Asn Ala His Pro Clu Phe Asn Gly
    CAC ATT CTT CCC CGC CTG ATC AAT GCT CAT CCC GAA TTC AAC GCC
```

M-75II

```
                    80                                      90
    Glu Val Cys Ser Clu Cln Ala Clu Thr Gly Pro Cys Arg Ala MET Ile Ser Arg Trp Tyr
    GAC CTG TGC TCT CAA CAA GCT CAC ACT GCC CCG TCC CCT GCA ATG ATC TCC CCC TGG TAC

                    100                                     110
    Phe Asp Val Thr Clu Gly Lys Cys Ala Pro Phe Phe Tyr Gly Gly Cys Gly Gly Asn Arg
    TTT CAT GTC ACT CAA CCT AAC TCC CCT CCA TTC TTT TAC CCC GGT TGC GCC GGC AAC CGT

                    120                                     130
    Asn Asn Phe Asp Thr Glu Clu Tyr Cys MET Ala Val Cys Cly Ser Ala Ile-COOH
    AAC AAC TTT GAC ACT GAA CAG TAC TCC ATG CCA GTG TGC CCC AGC GCT ATT
```

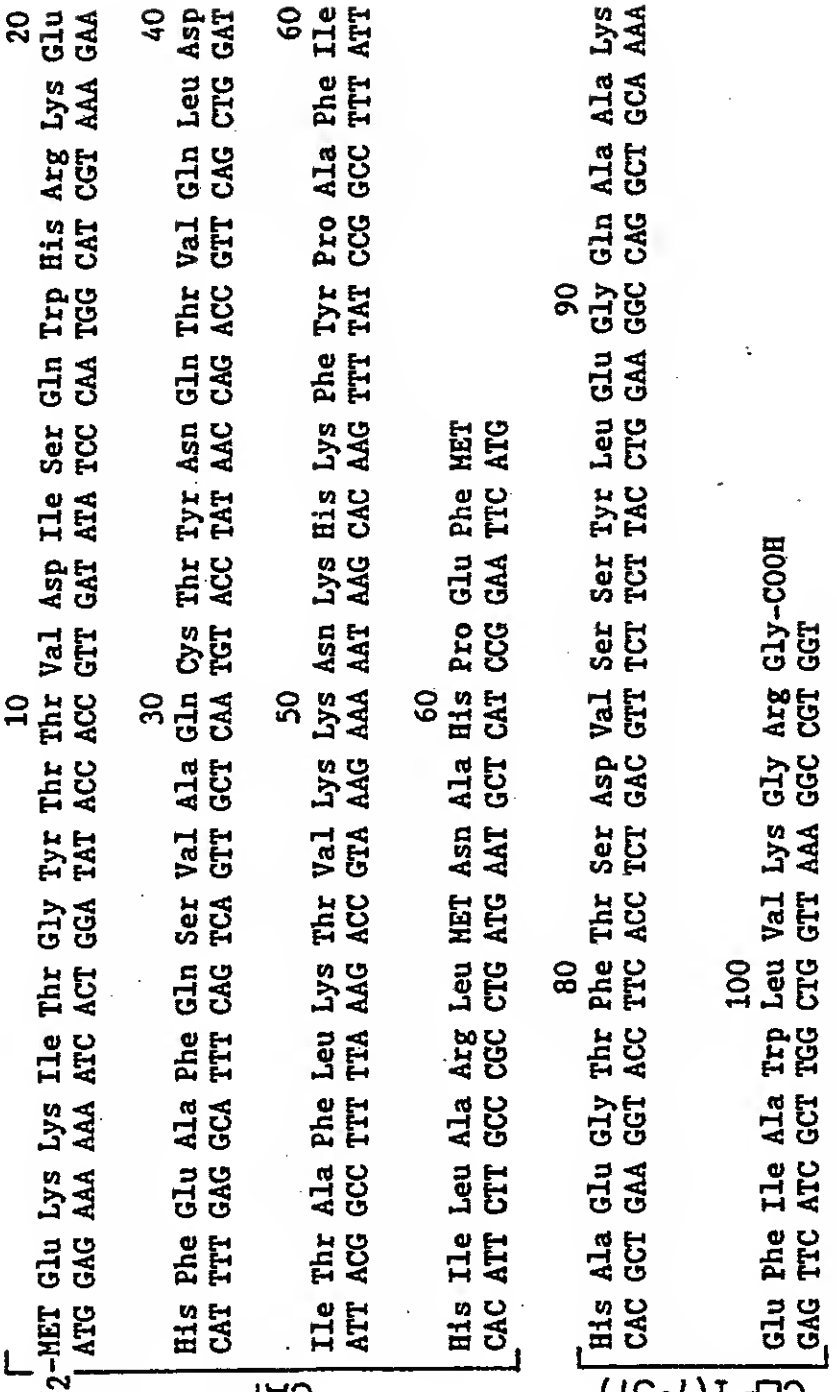
NH2-MET Glu Lys Lys Ile Thr Gly Tyr Thr Thr Val Asp Ile Ser Gln Trp His Arg Lys Glu
ATG GAG AAA AAA ATC ACT GGA TAT ACC ACC GTT GAT ATA TCC CAA TGG CAT CGT AAA GAA
10 20
His Phe Glu Ala Phe Gln Ser Val Ala Gln Cys Thr Tyr Asn Gln Thr Val Gln Leu Asp
CAT TTT GAG GCA TTT CAG TCA GTT GCT CAA TGT ACC TAT AAC CAG ACC GTT CAG CTG GAT
30 40
Ile Thr Ala Phe Leu Lys Thr Val Lys Lys Asn Lys His Lys Phe Tyr Pro Ala Phe Ile
ATT ACG GCC TTT TTA AAG ACC GTA AAG AAA AAT AAG CAC AAG TTT TAT CCG GCC TTT ATT
50 60
His Ile Leu Ala Arg Leu MET Asn Ala His Pro Glu Phe MET
CAC ATT CTT GCC CGC CTG ATG AAT GCT CAT CCG GAA TTC ATG
60
CAT
His Ala Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly Gln Ala Ala Lys
CAC GCT GAA GGT ACC TTC ACC TCT GAC GTT TCT TCT TAC CTG GAA GGC CAG GCT GCA AAA
80 90
Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly-COOH
GAG TTC ATC GCT TGG CTG GTT AAA GGC CGT GGT
100
CLP-I(7-37)

FIG. 4B

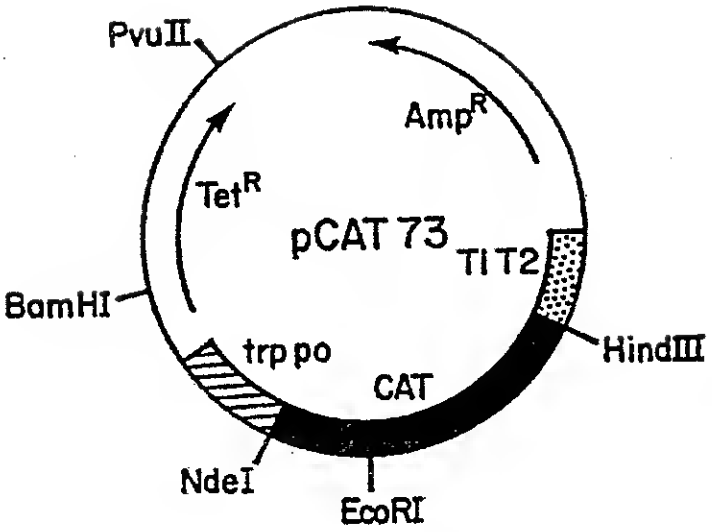
PvuII
Amp^R
Tet^R
pCAT 73
T1 T2
BamHI
HindIII
trp po
CAT
NdeI
EcoRI

FIG. 5A

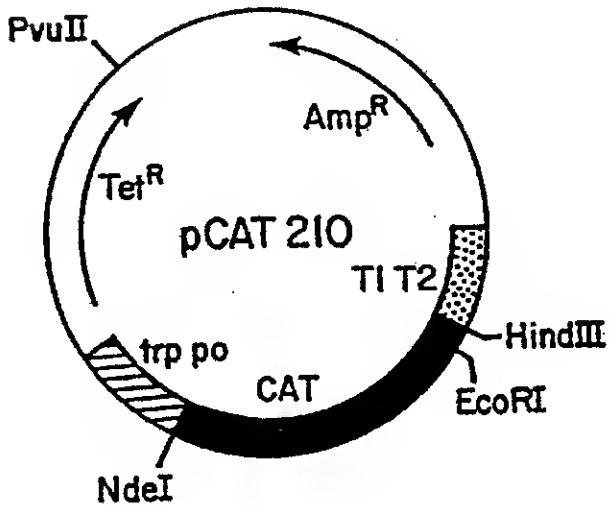
PvuII
Amp^R
Tet^R
pCAT 210
T1 T2
HindIII
trp po
CAT
EcoRI
NdeI

FIG. 5B

pC210SP-B

```
          GAATTCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGAACTAGTTAACTAGTACGCAAGTTCACGTAAAAAGGGTATCACAT

             100
ATG GAG AAA AAA ATC ACT GGA TAT ACC ACC GTT GAT ATA TCC CAA TGG CAT CGT AAA GAA CAT TTT GAG GCA TTT
MET Glu Lys Lys Ile Thr Gly Tyr Thr Thr Val Asp Ile Ser Gln Trp His Arg Lys Glu His Phe Glu Ala Phe

                                                 200
CAG TCA GTT GCT CAA TGT ACC TAT AAC CAG ACC GTT CAG CTG GAT ATT ACG GCC TTT TTA AAG ACC GTA AAG AAA
Gln Ser Val Ala Gln Cys Thr Tyr Asn Gln Thr Val Gln Leu Asp Ile Thr Ala Phe Leu Lys Thr Val Lys Lys

                                                                                 300
AAT AAG CAC AAG TTT TAT CCG GCC TTT ATT CAC ATT CTT GCC CGC CTG ATG AAT GCT CAT CCG GAA TTC CGT ATG
Asn Lys His Lys Phe Tyr Pro Ala Phe Ile His Ile Leu Ala Arg Leu MET Asn Ala His Pro Glu Phe Arg MET

GCA ATG AAA GAC GGT GAG CTG GTG ATA TGG GAT AGT GTT CAC CCT TGT TAC ACC GTT TTC CAT GAG CAA ACT GAA
Ala MET Lys Asp Gly Glu Leu Val Ile Trp Asp Ser Val His Pro Cys Tyr Thr Val Phe His Glu Gln Thr Glu
```

FIG. 6-I

```
                400
ACG TTT TCA TCG CTC TGG AGT GAA TAC CAC GAC GAT TTC CGG CAG TTT CTA CAC ATA TAT TCG CAA GAT GTG GCG
Thr Phe Ser Ser Leu Trp Ser Glu Tyr His Asp Asp Phe Arg Gln Phe Leu His Ile Tyr Ser Gln Asp Val Ala

                                        500
TGT TAC GGT GAA AAC CTG GCC TAT TTC CCT AAA GGG TTT ATT GAG AAT ATG TTT TTC GTC TCA GCC AAT CCC TGG
Cys Tyr Gly Glu Asn Leu Ala Tyr Phe Pro Lys Gly Phe Ile Glu Asn MET Phe Phe Val Ser Ala Asn Pro Trp

                                                                600
GTG AGT TTC ACC AGT TTT GAT TTA AAC GTG GCC AAT ATG GAC AAC TTC TTC GCC CCC GTT TTC ACC ATG GGC AAA
Val Ser Phe Thr Ser Phe Asp Leu Asn Val Ala Asn MET Asp Asn Phe Phe Ala Pro Val Phe Thr MET Gly Lys

TAT TAT ACG CAA GGC GAC AAG GTG CTG ATG CCG CTG GCG ATT CAG GTT CAT CAT GCC GTT TGT GAT GGC TTC CAT
Tyr Tyr Thr Gln Gly Asp Lys Val Leu MET Pro Leu Ala Ile Gln Val His His Ala Val Cys Asp Gly Phe His

                700                     CAT             linker          SP-B
GTC GGC AGA ATG CTT AAT GAA TTA CAA CAG|TCG GAT CCG GAA TTC AAC GGC|TTC CCC ATT CCT CTC CCC TAT TGC
Val Gly Arg MET Leu Asn Glu Leu Gln Gln|Ser Asp Pro Glu Phe Asn Gly|Phe Pro Ile Pro Leu Pro Tyr Cys
```

FIG. 6-2

```
                                                                800
TGG CTC TGC AGG GCT CTG ATC AAG CGG ATC CAA GCC ATG ATT CCC AAG GGT GCG CTA CGT GTG GCA GTG GCC CAG
Trp Leu Cys Arg Ala Leu Ile Lys Arg Ile Gln Ala MET Ile Pro Lys Gly Ala Leu Arg Val Ala Val Ala Gln

                                                                900
GTG TGC CGC GTG GTA CCT CTG GTG GCG GGC ATC TGC CAG TGC CTG GCT GAG CGC TAC TCC GTC ATC CTG CTC
Val Cys Arg Val Val Pro Leu Val Ala Gly Gly Ile Cys Gln Cys Leu Ala Glu Arg Tyr Ser Val Ile Leu Leu

                                                                974
GAC ACG CTG CTG GGC CGC ATG CTG CCC CAG CTG GTC TGC CGC CTC GTC CTC CGG TAA GCTT
Asp Thr Leu Leu Gly Arg MET Leu Pro Gln Leu Val Cys Arg Leu Val Leu Arg End
```

FIG. 6-3

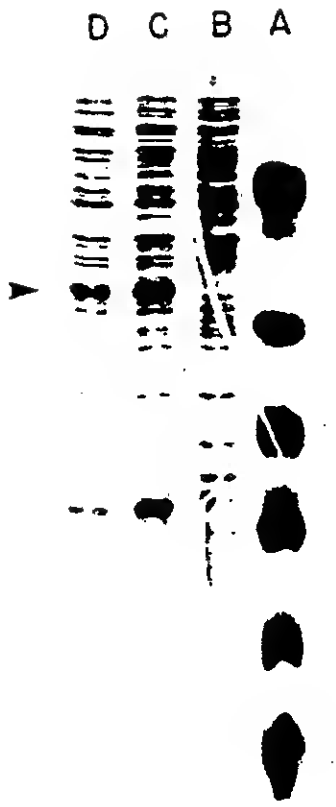

FIG. 7

FIG. 8-1

```
                pC210SP-C      GAATTCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGAA

                                                                  100
CTAGTTAACTAGTACGCAAGTTCACGTAAAAAGGGTATCACAT ATG GAG AAA AAA ATC ACT GGA
                                             MET Glu Lys Lys Ile Thr Gly

TAT ACC ACC GTT GAT ATA TCC CAA TGG CAT CGT AAA GAA CAT TTT GAG GCA TTT
Tyr Thr Thr Val Asp Ile Ser Gln Trp His Arg Lys Glu His Phe Glu Ala Phe

                                                 200
CAG TCA GTT GCT CAA TGT ACC TAT AAC CAG ACC GTT CAG CTG GAT ATT ACG GCC
Gln Ser Val Ala Gln Cys Thr Tyr Asn Gln Thr Val Gln Leu Asp Ile Thr Ala

TTT TTA AAG ACC GTA AAG AAA AAT AAG CAC AAG TTT TAT CCG GCC TTT ATT CAC
Phe Leu Lys Thr Val Lys Lys Asn Lys His Lys Phe Tyr Pro Ala Phe Ile His

                                         300
ATT CTT GCC CGC CTG ATG AAT GCT CAT CCG GAA TTC CGT ATG GCA ATG AAA GAC
Ile Leu Ala Arg Leu MET Asn Ala His Pro Glu Phe Arg MET Ala MET Lys Asp

GGT GAG CTG GTG ATA TGG GAT AGT GTT CAC CCT TGT TAC ACC GTT TTC CAT GAG
Gly Glu Leu Val Ile Trp Asp Ser Val His Pro Cys Tyr Thr Val Phe His Glu
```

FIG. 8-2

```
                                                400
CAA ACT GAA ACG TTT TCA TCG CTC TGG AGT GAA TAC CAC GAC GAT TTC CGG CAG
Gln Thr Glu Thr Phe Ser Ser Leu Trp Ser Glu Tyr His Asp Asp Phe Arg Gln

TTT CTA CAC ATA TAT TCG CAA GAT GTG GCG TGT TAC GGT GAA AAC CTG GCC TAT
Phe Leu His Ile Tyr Ser Gln Asp Val Ala Cys Tyr Gly Glu Asn Leu Ala Tyr

                    500
TTC CCT AAA GGG TTT ATT GAG AAT ATG TTT TTC GTC TCA GCC AAT CCC TGG GTG
Phe Pro Lys Gly Phe Ile Glu Asn MET Phe Phe Val Ser Ala Asn Pro Trp Val

AGT TTC ACC AGT TTT GAT TTA AAC GTG GCC AAT ATG GAC AAC TTC TTC GCC CCC
Ser Phe Thr Ser Phe Asp Leu Asn Val Ala Asn MET Asp Asn Phe Phe Ala Pro

        600
GTT TTC ACC ATG GGC AAA TAT TAT ACG CAA GGC GAC AAG GTG CTG ATG CCG CTG
Val Phe Thr MET Gly Lys Tyr Tyr Thr Gln Gly Asp Lys Val Leu MET Pro Leu

                                                700
GCG ATT CAG GTT CAT CAT GCC GTT TGT GAT GGC TTC CAT GTC GGC AGA ATG CTT
Ala Ile Gln Val His His Ala Val Cys Asp Gly Phe His Val Gly Arg MET Leu

                CAT                 linker          SP-C
AAT GAA TTA CAA CAG TCG GAT CCG GAA TTC AAC GGC ATT CCC TGC TGC CCA GTG
Asn Glu Leu Gln Gln Ser Asp Pro Glu Phe Asn Gly Ile Pro Cys Cys Pro Val
```

```
                                                                800
CAC CTG AAA CGC CTT CTT ATC GTG GTG GTG GTG GTC CTC ATC GTC GTG GTG
His Leu Lys Arg Leu Leu Ile Val Val Val Val Val Leu Ile Val Val Val

                                                848
ATT GTG GGA GCC CTG CTC ATG GGT CTC CAC TAA GCT T
Ile Val Gly Ala Leu Leu MET Gly Leu His End
```

FIG. 8-3

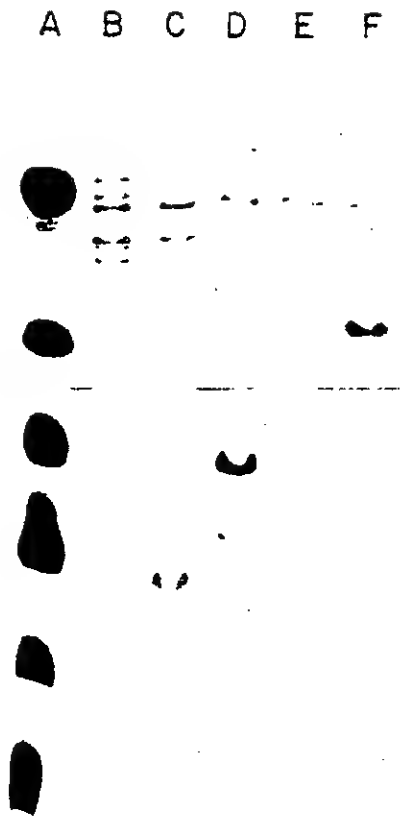

FIG. 9

FIG. 10-I

```
EcoRI                                 27                                      54
AAT TCG GGC GGG GCA GTT CTG GTC CTC CTA GGA GCG GCC GCC TGC GCC GCG CGG
Asn Ser Gly Gly Ala Val Leu Val Leu Leu Gly Ala Ala Ala Cys Ala Ala Arg

            BamHI                     81                                     108
CCC CGT GGT CGG ATC CTG GGC GGC AGA GAG GCC GAG GCG CAC GCT CGG CCT TAC
Pro Arg Gly Arg Ile Leu Gly Gly Arg Glu Ala Glu Ala His Ala Arg Pro Tyr

                                     135                                     162
ATG GCG TCG GTG CAG CTG AAC GGC GCG CAC CTG TGC GCA GGC GTC CTG GTG GCG
MET Ala Ser Val Gln Leu Asn Gly Ala His Leu Cys Ala Gly Val Leu Val Ala

                                     189                                     216
GAG CGG TGG GTG CTG AGC GCG GCG CAC TGC CTG GAG GAC GCG GCC GAC GGG AAG
Glu Arg Trp Val Leu Ser Ala Ala His Cys Leu Glu Asp Ala Ala Asp Gly Lys

                                     243                                     270
GTG CAG GTT CTC CTG GGC GCG CAC TCC CTG TCG CAG CCG GAG CCC TCC AAG CGC
Val Gln Val Leu Leu Gly Ala His Ser Leu Ser Gln Pro Glu Pro Ser Lys Arg

                                     297                                     324
CTG TAC GAC GTG CTC CGC GCA GTG CCC CAC CCG GAC AGC CAG CCC GAC ACC ATC
Leu Tyr Asp Val Leu Arg Ala Val Pro His Pro Asp Ser Gln Pro Asp Thr Ile
```

FIG. 10-2

```
                                        351                                              378
GAC CAC GAC CTC CTG CTG CTA CAG CTG TCG GAG AAG GCC ACA CTG GGC CCT GCT
Asp His Asp Leu Leu Leu Leu Gln Leu Ser Glu Lys Ala Thr Leu Gly Pro Ala

                                        405                                              432
GTG CGC CCC CTG CCC TGG CAG CGC GTG GAC CGC GAC GTG GCA CCG GGA ACT CTC
Val Arg Pro Leu Pro Trp Gln Arg Val Asp Arg Asp Val Ala Pro Gly Thr Leu

                                        459                                              486
TGC GAC GTG GCC GGC TGG GGC ATA GTC AAC CAC GCG GGC CGC CGC CCG GAC AGC
Cys Asp Val Ala Gly Trp Gly Ile Val Asn His Ala Gly Arg Arg Pro Asp Ser

                                        513                                              540
CTG CAG CAC GTG CTC TTG CCA GTG CTG GAC CGC GCC ACC TGC AAC CGG CGC ACG
Leu Gln His Val Leu Leu Pro Val Leu Asp Arg Ala Thr Cys Asn Arg Arg Thr

                                        567                                              594
CAC CAC GAC GGC GCC ATC ACC GAG CGC TTG ATG TGC GCG GAG AGC AAT CGC CGG
His His Asp Gly Ala Ile Thr Glu Arg Leu MET Cys Ala Glu Ser Asn Arg Arg

                                        621                                              648
GAC AGC TGC AAG GGT GAC TCC GGG GGC CCG CTG GTG TGC GGG GGC GTG CTC GAG
Asp Ser Cys Lys Gly Asp Ser Gly Gly Pro Leu Val Cys Gly Gly Val Leu Glu

                                        675                                              702
GGC GTG GTC ACC TCG GGC TCG CGC GTT TGC GGC AAC CGC AAG AAG CCC GGG ATC
Gly Val Val Thr Ser Gly Ser Arg Val Cys Gly Asn Arg Lys Lys Pro Gly Ile
```

```
                                    729                          StyI   756
TAC ACC CGC GTG GCG AGC TAT GCG GCC TGG ATC GAC AGC GTC CTG GCC TAG GGT
Tyr Thr Arg Val Ala Ser Tyr Ala Ala Trp Ile Asp Ser Val Leu Ala End

                                    783                                 810
GCC GGG GCC TGA AGG TCA GGG TCA CCC AAG CAA CAA AGT CCC GAG CAA TGA CCC

EcoRI                                       HindIII
GAA TTC TCA TGT TTG ACA GCT TAT CAT CGA TAA GCT T
```

FIG. 10-3

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US89/03417

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC94): C 12 N 1/20, 7/00, 15/00; C 12 P 21/00

II. FIELDS SEARCHED

Minimum Documentation Searched ?

Classification System	Classification Symbols
U.S.	435/68,252.33,235,320

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched *

CAS file 1967-1989, Biosis File 1967-1989

III. DOCUMENTS CONSIDERED TO BE RELEVANT *

Category *	Citation of Document, ** with indication, where appropriate, of the relevant passages **	Relevant to Claim No. **
Y	Gene, Volume 30, published 1984. P. Pumpen, et al. "Expression of hepatitis B virus surface antigen gene in Escherechia coli, pp. 201-210. see entire article.	1-15
Y	Science, Volume 237, published September 1987 K.S.Cook, et al. "Adipsin: a circulating serine protease homolog secreted by adipose tissue and sciatic nerve" pp. 402-405. see entire article	1,8,10
Y	J. Biol. Chem., published 25 July 1987. D.J. Drucker, et al. "Cell-specific post-translational processing of preproglucagon expressed from a metallothionein-glucagon fusion gene" pp. 9637-9643. see entire article.	1,8,10
Y	GB, A, 2173804 (Heynecker) 22 October 1986. see entire document	14,15

* Special categories of cited documents: **

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
26 November 1989	07 DEC 1989
International Searching Authority	**Signature of Authorized Officer**
ISA/US	Beth A. Burrous

Form PCT/ISA/210 (second sheet) (Rev. 11-87)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	Nucl. Acids. Res., Volume 15, number 9, published May 1987. "Expression of porcine pancreatic phospholipase A2. Generation of active enzyme by sequence-specific cleavage of a hybrid protein from Escherichia coli " pp. 3743-3759. see entire article.	6,7,9,12